Derivation of occupational exposure limits for airborne chemicals - Comparison of methods and protection levels

baua: Report

baua
Federal Institute for Occupational Safety and Health

Forschung
Projekt F 2437

Dr. K. Schneider
Dr. M. Dilger
Dr. E. Kaiser
Dr. F. Kalberlah
Dr. W. Wosniok

Derivation of occupational exposure limits for airborne chemicals - Comparison of methods and protection levels

1. Auflage 2022
Dortmund/Berlin/Dresden

Diese Veröffentlichung ist der Abschlussbericht zum Projekt „Derivation of occupational exposure limits for airborne chemicals –Comparison of methods and protection levels“ – Projekt F 2437 – im Auftrag der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Zitiervorschlag:
Dr. Klaus Schneider; Dr. Marco Dilger; Dr. Eva Kaiser; Dr. Fritz Kalberlah; Dr. Werner Wosniok, 2022. Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels. 1. Auflage. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin

Autoren: Dr. Klaus Schneider, Dr. Marco Dilger, Dr. Eva Kaiser, Dr. Fritz Kalberlah, Dr. Werner Wosniok
Forschungs- und Beratungsinstitut Gefahrstoffe GmbH
Klarastraße 63 • 79106 Freiburg

Fachliche Begleitung: Dr. Claudia Drossard, Heidi Ott, Prof. Dr. Thomas Gebel
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin

Titelfoto: peshkov/iStock.com

Umschlaggestaltung: Susanne Graul
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin

Herausgeber: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)
Friedrich-Henkel-Weg 1 – 25, 44149 Dortmund
Postanschrift: Postfach 17 02 02, 44061 Dortmund
Telefon 0231 9071-2071
Telefax 0231 9071-2070
E-Mail *info-zentrum@baua.bund.de*
Internet *www.baua.de*

doi:10.21934/baua:bericht20220112 (online)

www.baua.de/dok/8866360

Inhaltsverzeichnis

Ableitung von Luftgrenzwerten für chemische Stoffe am Arbeitsplatz - Vergleich von Methoden und Schutzniveaus

Kurzreferat

Die Ableitung und Festsetzung von Luftgrenzwerten für den Arbeitsplatz ist in verschiedenen nationalen und internationalen Verfahren ein wichtiger Bestandteil der Risikobewertung und des Risikomanagements von Chemikalien. Auf EU-Ebene ist eine Harmonisierung von Luftgrenzwerten aktuelles Thema, da sich für einzelne Stoffe aus den Verfahren im Rahmen des Arbeitsschutzrechts einerseits und des Chemikalienrechts andererseits unterschiedliche Grenzwerte für den Arbeitsplatz ergaben. Wesentliche methodische Aspekte der Ableitung von Arbeitsplatzgrenzwerten und analogen Werten sind die Bestimmung eines Startpunktes der Bewertung („Point of Departure“) anhand der in toxikologischen Studien beobachteten adversen Wirkungen sowie die Anwendung von Extrapolationsfaktoren zur Überbrückung von Datenlücken (bezüglich Unterschieden zwischen Studien mit unterschiedlicher Expositionsdauer, Unterschieden zwischen Versuchstier und Mensch, sowie bezüglich unterschiedlicher Empfindlichkeit zwischen den Menschen).

Dieses Projekt hatte zum Ziel, die Unterschiede zwischen den aktuell verwendeten Methoden der Grenzwertableitung und die resultierenden unterschiedlichen Schutzniveaus zu analysieren und transparent zu machen. In mehreren Teilprojekten wurden dazu

- die auf EU-Ebene sowie auf nationaler Ebene in Deutschland vorgeschlagenen und verwendeten Methoden analysiert und verglichen, und
- Datenauswertungen durchgeführt und darauf basierend Verteilungen für die zur Grenzwertableitung verwendeten Extrapolationsfaktoren erstellt.
- Mithilfe dieser Verteilungen und ihrer Verknüpfungen wurde analysiert, welchen Schutz vor nachteiligen Wirkungen die verwendeten Methodiken bieten und welches die wesentlichen Ursachen für Unterschiede sind.
- Das Projekt untersuchte in weiteren Teilprojekten wichtige Instrumente und Methoden der Grenzwertableitung
    - Dosis-Wirkungsmodellierung zur Bestimmung eines „Point of Departure“ mit dem Benchmark-Verfahren
    - Probabilistische Verfahren zur Beschreibung von Wahrscheinlichkeiten und Unsicherheiten bei der Ableitung von Arbeitsplatzgrenzwerten
    - Methoden zur kinetischen Modellierung von Aerosolen im Respirationstrakt zur Beschreibung von Speziesunterschieden und zur Bestimmung einer „Human Equivalent Concentration“ (HEC).

Übergeordnetes Ziel des Vorhabens war es, ein gemeinsames Verständnis für die notwendigen methodischen Festlegungen bei der Grenzwertsetzung und damit eine Grundlage zur Harmonisierung der Grenzwertableitung für den Arbeitsplatz in der EU zu schaffen.

Schlagwörter:

Arbeitsplatzgrenzwerte, Benchmark-Verfahren, Extrapolationsfaktoren, Unsicherheit, Zeitextrapolation, Interspeziesextrapolation, Intraspeziesextrapolation, Verteilungen, Probabilistische Risikobewertung, Schutzziel, humanäquivalente Konzentration

Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Abstract

The derivation and setting of occupational exposure limits (OELs) is an important component of the risk assessment and risk management of chemicals in different national as well as international processes. On the EU level the harmonisation of airborne exposure limit values is a current issue, because for some substances different exposure limits for workplaces were yielded by occupational safety and health legislation on the one hand and by chemicals legislation on the other hand. Important steps in the process of setting OELs or analogue values are the determination of a point of departure based on adverse effects reported in toxicological studies and the application of assessment factors to bridge data gaps (regarding studies with different exposure duration, differences between species and variability in sensitivity between humans).

The objective of this project was to analyse and disclose the differences between the current methods for deriving exposure limits and the resulting differences in protection levels. To achieve this,

- We analysed and compared the methodologies proposed or used at EU level and at national level in Germany
- Compiled data and used them for establishing distributions for assessment factors used for deriving exposure limits.
- With these distributions and their combinations, we analysed the level of protection achieved by the various methodologies and the reasons for differences.
- Further, we investigated important instruments and methods for deriving exposure limits
    - Dose-response modelling with the benchmark dose approach to determine the point of departure
    - Probabilistic approaches to describe probabilities and uncertainties of exposure limits
    - Methods for the modelling of kinetics of aerosols in the respiratory tract to describe respective interspecies differences and for determining a human equivalent concentration (HEC).

The overarching aim was to develop a common understanding of the necessary methodological steps in the setting of exposure limits and in this way to support harmonisation of the derivation of occupational exposure limits in the EU.

Keywords:

Occupational exposure limits (OELs), benchmark dose modelling, assessment factors (AFs), uncertainty, time extrapolation, interspecies extrapolation, intraspecies

extrapolation, distributions, probabilistic hazard assessment, protection level, human equivalent concentration

1 Einleitung

National und auf europäischer Ebene sind verschiedene Gremien und Institutionen mit der Ableitung von Arbeitsplatzgrenzwerten befasst. In Deutschland ist dies der Unterausschuss III des Ausschusses für Gefahrstoffe, unter Berücksichtigung der MAK-Werte der MAK-Kommission der Deutschen Forschungsgemeinschaft. In der EU legte das Scientific Committee on Occupational Exposure Limits (SCOEL) Empfehlungen für gesundheitsbasierte Grenzwerte vor, bis diese Aufgabe vom Ausschuss für Risikobewertung (Committee for Risk Assessment, RAC) der Europäischen Chemikalienagentur ECHA übernommen wurde. Mit dem Inkrafttreten von REACH gibt es darüber hinaus mit den „Derived No Effect Levels“ (DNELs) Beurteilungswerte für den Arbeitsplatz, die von den Stoffregistranten selbst abgeleitet werden. In Einzelfällen, z.B. für Stoffe, die der Zulassung unterliegen, gibt RAC darüber hinaus Empfehlungen für DNELs – unter anderem auch für den Arbeitsplatz – ab.

Wenn unterschiedliche Akteure Arbeitsplatzwerte für denselben Stoff ermitteln, sind numerische Unterschiede nicht ausgeschlossen. Dies kann bedingt sein durch Unterschiede in den verwendeten Daten und/oder ihrer Beurteilung, dem gewählten Schutzziel und der angewendeten Methodik zur Ermittlung der Werte. Unabhängig von der Ursache lösen alternative Werte Verunsicherung bei den Anwendern und ein Kommunikationsbedürfnis zur Erklärung der Unterschiede aus. Generell gilt der Anspruch an jede wissenschaftliche Bewertung, eine transparente und nachvollziehbare Begründung vorzulegen.

Vor diesem Hintergrund gibt es eine intensive und andauernde Diskussion um die bestehenden Methodiken zur Ableitung von Arbeitsplatzgrenzwerten und analoger Werte wie DNELs, deren Unterschiede und Weiterentwicklung (z.B. Deveau et al. 2015; Kalberlah und Heine 2015; Maier et al. 2015; Nies et al. 2013; Schenk 2010; Schenk et al. 2008; Schenk und Johanson 2011, 2018, 2019; Schenk et al. 2014). Ein wesentlicher methodischer Aspekt dabei ist die Verwendung von Extrapolationsfaktoren zur Überbrückung von Datenlücken. Während deren Verwendung bei Arbeitsplatzgrenzwerten eine eher neue Entwicklung ist (Dankovic et al. 2015; Schenk und Johanson 2018), wird zur Ableitung von gesundheitsbasierten Beurteilungswerten für die Allgemeinbevölkerung schon seit langem von Extrapolationsfaktoren Gebrauch gemacht (ECHA 2012; Falk-Filipsson et al. 2007; Kalberlah und Schneider 1998; Lehman und Fitzhugh 1954; WHO 1994). Wenn möglich, sollten Extrapolationsfaktoren substanzspezifisch ermittelt werden (Bhat et al. 2017) oder auf empirischen Daten zu anderen Substanzen beruhen (US EPA 2014). Die laufenden Bemühungen um eine Verbesserung der Bewertungsmethoden zielen unter anderem auf eine verbesserte wissenschaftliche Basis für die Quantifizierung der Extrapolationsfaktoren und die Einführung neuer Methoden wie das Benchmark-Verfahren ab.

Dieses Projekt verfolgte das Ziel, die wissenschaftlichen Grundlagen der Ableitung von Grenzwerten für den Arbeitsplatz zu analysieren und zu aktualisieren, um so zu einer Harmonisierung der Methoden beizutragen.

Dazu wurden

- bestehende methodische Systeme analysiert und verglichen
- neue methodische Ansätze wie z.B. das Benchmark-Verfahren und probabilistische Verfahren mit ihren Vor- und Nachteilen dargestellt
- auf Basis aktueller Datenauswertungen Vorschläge für geeignete Verteilungen für Extrapolationsfaktoren gemacht
- und der durch die unterschiedlichen Ableitungsmethoden erreichte Schutz transparent gemacht.

2 Überblick über die im Projekt erarbeiteten Berichte

Die im Projekt erarbeiteten Ergebnisse wurden in 10 Einzelberichten zusammengefasst, die in diesen Endbericht integriert sind. Nachfolgend wird eine kurze Übersicht zu deren Inhalten gegeben.

Report 1: Comparison of methods for deriving OELs

In diesem Bericht werden die Methodiken zur Ableitung von Arbeitsplatzgrenzwerten oder analogen Werten in Deutschland und auf EU-Ebene analysiert und verglichen, schwerpunktmäßig bezüglich

- der Definition und des Gültigkeitsbereichs der Werte
- der Datensuche und -auswertung
- der Methodik zur Ableitung von Werten für systemische Effekte
- der Methodik zur Ableitung von Werten für lokale Wirkungen im Atemtrakt.

Die Analyse zeigt, dass Unterschiede aus allen methodischen Schritten des Ableitungsweges resultieren können, und führte zu Empfehlungen für die Harmonisierung der Methoden sowie für die Verbesserung der Transparenz durch detailliertere Leitfäden.

Report 2: Benchmark dose modelling

Dieser Bericht beschreibt den aktuellen Stand der Anwendung des Benchmark-Verfahrens in der Risikobewertung sowie die verfügbaren methodischen Beschreibungen und Instrumente (Software). Auf Besonderheiten wie die Bestimmung der „benchmark response“ im Falle von quantalen und kontinuierlichen Daten wird eingegangen.

Report 3: Benchmark dose modelling - Examples

Zur Erläuterung des Benchmark-Verfahrens wurden als Beispiele Modellierungen für jeweils 5 Stoffe mit quantalen bzw. kontinuierlichen Daten durchgeführt, die in diesem Bericht dargestellt sind.

Report 4: Probabilistic hazard assessment

Die Grundsätze der probabilistischen Risikobewertung werden erläutert: Der Point of Departure und die Extrapolationsfaktoren werden dabei als Verteilungen dargestellt und mit Monte-Carlo-Verfahren verknüpft. Der Bericht beschreibt weiterhin verfügbare Instrumente (Software), demonstriert die Anwendbarkeit des probabilistischen Verfahrens anhand von 2 Beispielen und diskutiert die Möglichkeiten des Verfahrens, Unsicherheiten in der Risikobewertung zu beschreiben.

Report 5: Route-to-route extrapolation

In einem knappen, zusammenfassenden Bericht wird beschrieben, unter welchen Bedingungen eine Extrapolation von einem Expositionspfad zu einem anderen möglich ist, und wie sie durchgeführt werden kann.

Report 6: Time extrapolation

Zur Erstellung von Verteilungen für die Zeitextrapolation (subakut zu chronisch, subakut zu subchronisch, subchronisch zu chronisch) wurden zwei Datenquellen ausgewertet: Studien des US National Toxicology Program sowie Daten aus der ECHA-Datenbank der REACH-Registrierungsdossiers. Letztere wurden von der ECHA in einer anonymisierten, standardisierten Form überlassen und durch eine im Projekt erstellte semi-automatische Routine ausgewertet. Die Auswertung der NTP-Daten erfolgte händisch. Die erhaltenen empirischen Datensätze zur Zeitextrapolation wurden mit publizierten Ergebnissen anderer Autoren verglichen.

Report 7: Interspecies extrapolation

Die unter “Time extrapolation“ beschriebenen Datensätze von NTP-Studien und REACH-Daten wurden auch hinsichtlich der quantitativen Unterschiede zwischen verschiedenen Versuchstierspezies ausgewertet. Diese Auswertung ergab eine gute Übereinstimmung mit Vorhersagen des allometrischen Scalings. Die erhaltenen empirischen Datensätze zur Interspeziesextrapolation wurden wiederum mit publizierten Ergebnissen anderer Autoren verglichen.

Report 8: Intraspecies extrapolation

Zur Verbesserung der Datenlage wurden Veröffentlichungen zu toxikokinetischen Unterschieden beim Menschen zusammengestellt und quantitativ ausgewertet. Eine weitere Auswertung von publizierten Humandaten betraf (toxikodynamische) Unterschiede in individuellen Dosen, die zu ähnlichen Effektstärken führten. Weiterhin wurden die von Abdo et al. (2015) publizierten in vitro-Daten ausgewertet. Die eigene Datenauswertung zu toxikokinetischen Unterschieden sowie der Datensatz von Abdo et al. (2015) zu toxikodynamischen Unterschieden wurden als geeignet angesehen, um Verteilungen zur Intraspeziesextrapolation zu bilden.

Report 9: Human equivalent concentration and kinetic modelling of aerosols in the lower respiratory tract

Deposition und Clearance von Aerosolen in Versuchstieren und Menschen können modelliert und so toxikokinetische Unterschiede zwischen den Spezies beschrieben werden. Dieser Bericht erläutert die Prozeduren und wie daraus eine

humanäquivalente Konzentration (human equivalent concentration, HEC) errechnet werden kann und diskutiert die inhärenten Unsicherheiten der Methodik.

Report 10: Synthesis report: Modelling of distributions of assessment factors, comparison with current methods and discussion of protection goals

Im letzten Schritt wurden die in diesem Projekt für die einzelnen Extrapolationsschritte parametrisierten Verteilungen mittels probabilistischer Monte-Carlo-Verfahren kombiniert. Durch Vergleich mit derzeit für die Grenzwertableitung verwendeten Extrapolationsfaktoren konnte die Wahrscheinlichkeit beschrieben werden, mit der die verschiedenen Methodiken hinreichend Schutz vor adversen Wirkungen bieten. Zwei probabilistisch behandelte Bewertungsbeispiele weiteten die Betrachtung auf den Point of Departure aus. Aus den beobachteten Unterschieden wurden Vorschläge zur Harmonisierung der Methodiken abgeleitet.

3 Ausblick

Die Ergebnisse dieses Projektes dienen dazu, die Methodik der Grenzwertableitung für chemische Stoffe am Arbeitsplatz in der EU zu vereinheitlichen. Der unmittelbar erwartete Nutzen besteht in der Schaffung einer wissenschaftlichen Diskussionsgrundlage, auf deren Basis die internationale Diskussion befördert und die Harmonisierung vorangetrieben werden kann.

Hierzu ist geplant, die Projektergebnisse im Rahmen eines internationalen Workshops im Frühjahr 2022 in Dortmund vorzustellen. Weiterhin wurden zwei wissenschaftliche Veröffentlichungen erarbeitet, die die Projektergebnisse zusammenfassen und die beim Journal of Applied Toxicology eingereicht werden.

Aus einer übergeordneten Sicht dienen die Projektergebnisse dazu

- die Transparenz und Verlässlichkeit der Bewertungen von Gremien und anderen Akteuren zu steigern und dadurch die Wahrnehmung dieser Bewertungen als verlässliche Grundlage des Risikomanagements zu verbessern
- die Unsicherheit im Umgang mit den Grenzwerten, zum Beispiel im betrieblichen Alltag bei der Durchführung von Gefährdungsbeurteilungen und Erarbeitung von Arbeitsanweisungen, zu verringern
- und damit den Gesundheitsschutz der Arbeitnehmer zu verbessern.

Zusammenfassung

Arbeitsplatzgrenzwerte sind wichtige Instrumente zur Kontrolle und Beherrschung der Exposition gegenüber Gefahrstoffen am Arbeitsplatz. Verschiedene nationale und internationale Gremien sind mit der Ableitung von Arbeitsplatzgrenzwerten befasst. In der Europäischen Union erarbeitete der Wissenschaftliche Ausschuss für Grenzwerte berufsbedingter Exposition (Scientific Committee on Occupational Exposure Limits,SCOEL) Vorschläge für solche Werte, bis der Ausschuss für Risikobewertung (Committee for Risk Assessment, RAC) der Europäischen Chemikalienagentur (ECHA) diese Aufgabe 2019 übernahm. In Deutschland beschäftigen sich zwei Gremien mit der Ableitung von gesundheitsbasierten Arbeitsplatzgrenzwerten: der Ausschuss für Gefahrstoffe (AGS) und die Ständige Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe (MAK-Kommission) der Deutschen Forschungsgemeinschaft (DFG). Gesundheitsbasierte Beurteilungswerte für den Arbeitsplatz werden jedoch nicht nur im Rahmen der Arbeitsschutzgesetzgebung, sondern auch in anderen Regulationsbereichen aufgestellt. Im Bereich der EU-Chemikalienverordnung REACH (EC 2006) sind „Derived No Effect Levels“ (DNELs) Bestandteil der Stoffsicherheitsbewertung in Registrierungsdossiers, die von den einreichenden Firmen erstellt werden. “Acceptable Exposure Levels” (AELs) und “Acceptable Operator Exposure Levels” (AOELs) für Wirkstoffe in Biozid- und Pflanzenschutzprodukten werden nach Maßgaben der EU Biozidprodukteverordnung (EU 2021) und der EU Pflanzenschutzprodukterichtlinie (EC 2021) abgeleitet. Solche Werte werden nachfolgend als Arbeitsplatzgrenzwert-analoge Werte zusammengefasst.

Wir analysierten und diskutierten die **Methoden** verschiedener Systeme zur Ableitung von Arbeitsplatzgrenzwerten und analoger Werte mit dem Ziel, Unterschiede zu identifizieren und Empfehlungen für eine Harmonisierung zu erarbeiten. Dazu untersuchten wir schwerpunktmäßig folgende Aspekte: Definition und Gültigkeitsbereich der Werte, Datensuche und -auswertung, Methodik zur Ableitung von Werten für systemische bzw. lokale Effekte im Atemtrakt.
Die einzelnen betrachteten Unterpunkte wurden für die verschiedenen Systeme übersichtlich in Tabellen gegenübergestellt (Report 1, Tabellen 2-2 – 2-5).
Die Analyse zeigte auf, dass auf allen Stufen des Ableitungsprozesses quantitative Unterschiede auftreten können, insbesondere bei der Bestimmung des Startpunktes der Bewertung („Point of departure“, POD) und bei der Anwendung von Extrapolationsfaktoren und entsprechender Standardwerte. Empfehlungen zur Harmonisierung der Ableitungssysteme betreffen insbesondere folgende Aspekte:

- Die Verwendung der “Benchmark dose-Methodik” zur Bestimmung des POD
- Die Verwendung des allometrischen Scalings
- Die Höhe der verwendeten Extrapolationsfaktoren (z.B. für Zeit- und Intraspeziesextrapolation)
- Die Notwendigkeit und Anwendung von endpunktspezifischen Extrapolationsfaktoren (z.B. für Reproduktionstoxizität, Atemwegstoxizität und sensorische Reizung).

Unzureichende oder fehlende Leitfäden zur Ableitung von Arbeitsplatzgrenzwerten wurden als ein Grund für mangelnde Transparenz und als ein Hindernis für weitergehende Harmonisierung identifiziert. Wir erachten deswegen eine komplette und detaillierte Dokumentation der Methodiken für wichtig, insbesondere bezüglich:

- der Auswahl eines geeigneten POD (z.B. bei Vorliegen verschiedener geeigneter Studien oder bei der Abwägung der Beweiskraft von Daten)
- der Modifizierung des POD, um Unterschieden zwischen Experiment und den Expositionsbedingungen am Arbeitsplatz Rechnung zu tragen
- der Anwendung der einzelnen Extrapolationsfaktoren und der Bedingungen, unter denen von Standardannahmen abgewichen werden kann.

In einem weiteren Bericht wird eine aktuelle Übersicht zur Anwendung der **Benchmark dose (BMD)-Modellierung** in der Risikobewertung gegeben. Mit speziellen Instrumenten (Programmen oder Online-Anwendungen) werden dabei mathematische Modelle an experimentelle Daten angepasst, sodass die Modelle die Dosis-Wirkungsbeziehung adäquat beschreiben. Die kritische Effektstärke wird als „benchmark response" (BMR) bezeichnet. Die der BMR entsprechende Dosis wird "benchmark dose" (BMD) genannt und benennt die Dosis, bei der der durch die BMR definierte zusätzliche Effekt (gegenüber der Kontrolle) erwartet wird. Vertrauensintervalle zur BMD beschreiben deren Unsicherheit, die aus den experimentellen Bedingungen, Messunsicherheiten etc. resultieren. Die BMDL („benchmark dose lower bound") beschreibt die Untergrenze eines einseitigen 95%-Vertrauensintervalls der BMD, die BMDU die Obergrenze eines einseitigen 95%-Vertrauensintervalls. Sowohl BMD als auch BMDL werden als POD verwendet, jedoch ist die BMDL zu bevorzugen, da sie die Unsicherheit der BMD berücksichtigt. Es existieren Leitfäden zur Anwendung des Benchmark-Verfahrens der Europäischen Lebensmittelsicherheitsagentur (EFSA), der US Environmental Protection Agency (US EPA) und der Weltgesundheitsorganisation (Environmental Health Criteria Document 240, Chapter 5). Neuere Weiterentwicklungen der auf der Software PROAST des National Institute for Public Health and the Environment (RIVM) bzw. der Software BMDS der US EPA basierenden Anwendungen ermöglichen es, die Modellunsicherheit durch sogenannte Model averaging-Verfahren zu berücksichtigen. Es wird deswegen empfohlen, Model averaging für quantale und kontinuierliche Daten anzuwenden.
In den existierenden Leitfäden zur Ableitung von Arbeitsplatzgrenzwerten wird das Benchmark-Verfahren meist nur als Alternative zur Extrapolation von einem Lowest observed adverse effect level (LOAEL) auf einen No adverse effect level (NAEL) bei fehlendem No observed adverse effect level (NOAEL) beschrieben. Hintergrundinformationen und Anleitung zur Anwendung des Verfahrens sind häufig unzureichend. Dies steht nicht im Einklang mit der Einschätzung, dass dieses Verfahren das derzeit beste zur Bestimmung des POD darstellt.

Die Anwendung des **Benchmark-Verfahrens** wurde anhand von 10 **Beispielen** in einem zusätzlichen Bericht demonstriert. Es wurden fünf Beispielsubstanzen mit quantalen und fünf mit kontinuierlichen Daten ausgesucht. Diese Beispiele beinhalten auch Datensätze, bei denen kein NOAEL, sondern nur ein LOAEL bestimmt werden konnte, sowie ein Beispiel mit epidemiologischen Daten.

Die Anwendbarkeit von **probabilistischen Verfahren** in der Risikobewertung war Gegenstand eines weiteren Teilberichts. In probabilistischen Verfahren werden die Eingangsdaten (die BMD, Extrapolationsfaktoren) der Grenzwertableitung als Verteilungen dargestellt, die die Unsicherheit der BMD und der Extrapolationsfaktoren sowie die Variabilität innerhalb der Zielpopulation beschreiben. Diese Verteilungen werden mittels probabilistischer Methoden (Monte-Carlo-Simulationen) kombiniert, womit eine Verteilung des Grenzwertes erhalten wird. Dies erfordert Festlegungen bezüglich der Effektstärke (BMR) zur Ableitung der BMD, bezüglich des Anteils der Zielpopulation, der als Schutzziel einbezogen werden soll und bezüglich der Wahrscheinlichkeit für das Erreichen des intendierten Schutzziels.
Die erhaltene Grenzwertverteilung erlaubt es, die Unsicherheit und die Variabilität des Ergebnisses zu beschreiben, das Niveau des erreichten Schutzes zu charakterisieren ebenso wie die Wahrscheinlichkeit und Stärke von Effekten bei höheren Konzentrationen.
Zwei auch für Nicht-Statistiker einfach zu verwendende Instrumente stehen zur Verfügung:

- Das APROBA-Tool, das im Rahmen des WHO/IPCS-Projektes “Evaluating and Expressing Uncertainty in Hazard Characterization” entwickelt wurde (WHO 2014).
- Das Monte-Carlo-Tool der EFSA.

Aufgrund dieser neuen Entwicklungen sind probabilistische Verfahren einfacher anzuwenden. Wir empfehlen ihren Einsatz für die Entwicklung und Diskussion der Methoden zur Ableitung gesundheitsbasierter Grenzwerte, zur Festlegung deterministischer Werte für Extrapolationsfaktoren (und deren Kombinationen) und zum Vergleich mit deterministischen Stoffbewertungen sowie für komplexe Einzelfälle.

Ein kurzer Bericht fasst die Möglichkeiten und Limitierungen der **Pfad-zu-Pfad-Extrapolation** zusammen. Diese kann unter bestimmten Bedingungen zur Ableitung von Luftgrenzwerten für den Arbeitsplatz angewendet werden, wenn keine belastbaren Inhalationsstudien, wohl aber Studien für einen anderen Expositionspfad vorliegen. Von der Interdepartmental Group on Health Risks from Chemicals (IGHRC) sowie von Geraets et al. (2014) wurden Kriterien und Bedingungen für die Pfad-zu-Pfad-Extrapolation formuliert. Eine schrittweise praktische Anleitung bietet der Leitfaden der ECHA (“Guidance document on Information Requirements and Chemical Safety Assessment, R.8”) (ECHA 2012). Generell wird eine Pfad-zu-Pfad-Extrapolation nur befürwortet, wenn die kritischen Effekte systemischer Natur sind und keine pfad-spezifischen Unterschiede existieren, die eine Vorhersage für den anderen Pfad unsicher machen, wie z.B. eine ausgeprägte Metabolisierung in der Leber vor Eintritt in den systemischen Kreislauf („first-pass effect“).

Ein zentraler Bestandteil des Projektes bestand in der Analyse der Extrapolationsschritte, die zur Ableitung von Arbeitsplatzgrenzwerten angewendet werden (Zeit-, Inter- und Intraspeziesextrapolation) und in der Verbesserung der empirischen Datenlage zu ihrer Ableitung.
Im Falle der **Zeit- und Interspeziesextrapolation** wurden Verteilungen für die Verhältnisse von NOAEL (oder LOAEL-Werten oder Vergleichbarem) aus Studien unterschiedlicher Expositionsdauer bzw. unterschiedlicher Spezies abgeleitet. Dazu wurden Daten aus Studienberichten des US National Toxicology Program (NTP) sowie

Daten aus REACH-Registrierungen ausgewertet. Daten aus NTP-Studien wurden manuell ausgewertet: Aus jedem Bericht wurden Dosen oder Konzentrationen, die dem NOAEL und/oder LOAEL entsprachen, für jeden Studientyp (pro Spezies, Geschlecht und Studiendauer) sowie für jede ausgewertete Art von Endpunkt (Körpergewicht, lokale oder systemische Wirkungen) extrahiert. Im Falle der REACH-Daten wurden solche Werte aus strukturierten Daten ermittelt, die von der ECHA aus ihrer Datenbank der Registrierungsdaten extrahiert und zur Verfügung gestellt wurden. Die REACH-Daten beinhalteten wesentlich mehr toxikologische Studien mit wiederholter Applikation als die NTP-Studien. Jedoch mussten strikte Auswahlkriterien angewendet werden, um eine hinreichende Qualität des Datensatzes zu gewährleisten.

Sowohl für die NTP- als auch für die REACH-Daten wurden substanzspezifisch alle verfügbaren Studienpaare für die Paarungen subakut/chronisch, subchronisch/chronisch und subakut/subchronisch verglichen und Verhältniswerte für die Paarungen $Dosis_{kürzere\ Studie}/Dosis_{längere\ Studie}$ berechnet. Die Verteilungen dieser Verhältniswerte wurden nach verschiedenen Parametern abgeschichtet, um deren Einfluss zu testen. Diese Prüfungen ergaben für beide Datensätze, dass keine konsistenten Einflüsse auf die Verteilungen zur **Zeitextrapolation** erkennbar waren bezüglich folgender Parameter:

- Expositionspfad (oral, Inhalation)
- Geschlecht
- Spezies
- Lokale versus systemische Effekte nach Inhalation
- Zielorgane
- Substanzgruppen (exemplarisch untersucht anhand von zwei Substanzgruppen untersucht in NTP-Studien).

Trotz strenger Auswahlkriterien wurden Qualitätsprobleme bei den REACH-Daten sichtbar. Außerdem führte die semi-automatische Prozessierung der REACH-Daten, die aufgrund der großen anfänglichen Studienzahl notwendig war, zu Nachteilen in der Interpretierbarkeit gegenüber der manuellen Auswertung der NTP-Daten. Aus diesen Gründen beurteilten wir die REACH-Daten als weniger verlässlich und aussagekräftig als die NTP-Daten. Diese Schlussfolgerung ist im Einklang mit einer für die REACH-Daten beobachteten größeren geometrischen Standardabweichung (GSD), die auf eine höhere Streuung in diesen Daten hinweist. Deswegen wurde der kombinierte Datensatz von oralen und inhalativen NTP-Studien zur Ableitung von Verteilungen für die Zeitextrapolation als geeignet angesehen. Aus NTP-Studien von 256 Stoffen wurden ca. 400 Verhältniswerte jeweils für subakut/chronisch und subakut/subchronisch sowie mehr als 1200 Verhältniswerte für die Paarung subchronisch/chronisch ermittelt. Für die Zeitextrapolation im Falle von lokalen Wirkungen im Atemtrakt werden dieselben Verteilungen zur Verwendung vorgeschlagen wie im Falle von systemischen Effekten, da diesbezüglich keine Unterschiede gefunden wurden. Diese Beobachtung wird durch andere Publikationen gestützt, die ähnliche oder höhere Verhältniswerte für lokale Wirkungen fanden als für systemische Wirkungen.
Unsere Analyse ist eine der wenigen Auswertungen, die alle Schritte der Zeitextrapolation beinhaltet (subakut/subchronisch, subchronisch/chronisch,

subakut/chronisch). Die geometrischen Mittelwerte (GM) der Verhältniswerte aus der Auswertung der NTP-Studien für die Schritte subakut/chronisch (GM 4,11) und subakut/subchronisch (GM 1,60) passen gut zu anderen publizierten Auswertungen der vergangenen Jahre. Der GM der Verteilung subchronisch/chronisch (GM 2,93) liegt am oberen Ende der von anderen berichteten Werte. Der durch Multiplikation unserer GM der zwei Teilschritte erhaltene Wert stimmt jedoch gut mit dem GM für subakut-chronisch überein, was für eine hohe Konsistenz der Ergebnisse spricht.

Die Extrapolation von experimentellen Tierstudien auf den Menschen (**Interspeziesextrapolation**) ist ein entscheidender Schritt bei der Ableitung vieler Arbeitsplatzgrenzwerte oder analoger Werte. Die bestehenden Methoden verwenden entweder einen Standardfaktor (der oft in einen toxikokinetischen und einen toxikodynamischen Teilschritt unterschieden wird) oder verwenden allometrische Scalingfaktoren. Allometrisches Scaling beschreibt eine Korrelation von physiologischen und kinetischen Parametern mit dem Körpergewicht, potenziert mit einem Exponenten. Für ein Scaling nach kalorischem Grundumsatz ist der allometrische Exponent 0,75. Sowohl NTP-Studien als auch REACH-Daten wurden auch bezüglich Speziesunterschieden ausgewertet. Substanz- und studienspezifische Dosiswerte wurden ausgewertet wie oben für die Zeitextrapolation beschrieben. In Übereinstimmung mit den Erwartungswerten des allometrischen Scalings zeigten die Auswertungen der oralen Studien aus beiden Datensätzen, dass die größere Spezies sensibler erscheint, wenn Dosisangaben als Substanzmenge pro kg Körpergewicht verglichen werden (GM für die Verhältniswerte Ratte/Maus: NTP: 0,40; REACH: 0,66; Erwartungswert: 0,59).
Wiederum in Übereinstimmung mit allometrischen Grundsätzen zeigten die Inhalationsstudien mit Ratten und Mäusen (für die die meisten Daten verfügbar waren) eine ähnliche Empfindlichkeit der Spezies (GM für die Verhältniswerte Ratte/Maus: NTP: 0,96; REACH: 1,09; Erwartungswert: 1) für den Fall, dass Expositionskonzentrationen verglichen wurden. In allen Datensätzen wurde eine relevante Variabilität um die GM beobachtet (GSD: NTP: 3,6-3,8; REACH: 3,0-3,9).
Diese Ergebnisse wurden mit anderen Auswertungen zu Speziesunterschieden bezüglich Toxizität und toxikokinetischer Daten verglichen. Obwohl die enthaltene Unsicherheit keinen Rückschluss erlaubt, ob ein allometrischer Exponent von 0,7 oder 0,75 geeigneter erscheint, unterstreicht die Auswertung die Anwendbarkeit des allometrischen Scalings. Ein Scaling nach kalorischem Grundumsatz wird empfohlen, da es sowohl durch empirische Daten als auch durch mechanistische Überlegungen gestützt wird. Dies führt auch zur Schlussfolgerung, dass keine Korrektur durch Scalingfaktoren notwendig ist, wenn ein Grenzwert ausgehend von einer Luftkonzentration als POD abgeleitet wird. Basierend auf diesen Datenauswertungen wird ein zweistufiges Vorgehen empfohlen:

- Korrektur (“Normalisierung”) der Dosen durch allometrische Scalingfaktoren (nach kalorischem Grundumsatz) für orale Daten (für Konzentrationen aus Inhalationsstudien ist keine Korrektur notwendig)
- Berücksichtigung der verbleibenden Unsicherheit (aufgrund der Unterschiede von Substanz zu Substanz) durch eine Verteilung, die aus der Auswertung der Daten erhalten wird.

Aufgrund der höheren Qualität der NTP-Datenauswertung wird dieser der Vorzug gegenüber den REACH-Daten gegeben.

Zur Beschreibung von Speziesunterschieden aufgrund von Unterschieden in der Deposition und Clearance von Aerosolen im Atemtrakt existieren spezifische Modellierungsansätze. Das Verfahren der humanäquivalenten Konzentration ("**Human Equivalent Concentration**", HEC) zielt darauf ab, aus einer Expositionskonzentration einer tierexperimentellen Inhalationsstudie eine äquivalente (d.h. äquipotente) Konzentration für den Menschen (Arbeitsplatzszenario) zu berechnen. In dem entsprechenden Teilbericht diskutieren wir die Schritte der HEC-Berechnung und die Möglichkeiten, Deposition und Clearance von Partikeln im unteren Atemtrakt mit dem "Multiple Pathway Deposition Model" (MPPD) zu berechnen. Es wurde eine erhebliche Unsicherheit bei der Quantifizierung der Speziesunterschiede bezüglich Deposition und Elimination sichtbar. Einer der Hauptursachen der Unsicherheit ist die fehlende Kenntnis zu geeigneten Einheiten für die Beschreibung der deponierten Stoffmenge (Normalisierung), zum Beispiel als Menge pro Lungenoberfläche oder pro Volumeneinheit der Alveolarmakrophagen. Angesichts der hohen Unsicherheit des HEC-Verfahrens schlagen wir vor, die Unsicherheit mit derselben Verteilung zu berücksichtigen wie für die Interspeziesextrapolation bei systemischen Wirkungen.

Die Beschreibung von Empfindlichkeitsunterschieden in der menschlichen Bevölkerung (**Intraspeziesextrapolation**) ist ein weiterer zentraler Aspekt der Grenzwertableitung. Die innerartliche Variabilität der Empfindlichkeit kann in toxikokinetischen (das bedeutet unterschiedliche innere Exposition bei gleicher äußerer Exposition) oder toxikodynamischen Unterschieden (unterschiedliche Wirkung im Zielgewebe bei gleicher innerer Exposition) begründet sein. Verschiedene Bedingungen und Parameter können solche Unterschiede beeinflussen oder verursachen, z.B. Alter, Geschlecht, genetische Faktoren (z.B. Polymorphismen in fremdstoffmetabolisierenden Enzymen), epigenetische Unterschiede, sowie gesundheitliche Einflüsse. Entsprechend ist die Quantifizierung der innerartlichen Variabilität der Empfindlichkeit eine Herausforderung. Gegenwärtig für die Ableitung von Arbeitsplatzgrenzwerten benutzte Extrapolationsfaktoren sind ungenügend datenbasiert. Als Teil dieses Projektes wurden Daten zur innerartlichen Variabilität aus publizierten Humanstudien recherchiert, zusammengestellt und ausgewertet:

- 78 Studien (68 davon quantitativ auswertbar) zu toxikokinetischen Unterschieden
- 25 Studien zu toxikodynamischen Unterschieden.

Die Variabilität bezüglich **toxikokinetischer** Daten wurde durch log GSD-Werte charakterisiert (der Standardabweichung der logarithmierten Daten). Der Median der log GSD-Werte aller Daten lag bei 0,146, was einem Faktor von ca. 1,7 zwischen dem Median und dem 95. Perzentil der Population entspricht. Das 95. Perzentil des log GSD von 0,355 entspricht einem Faktor von 3,8, der somit 95% der Zielpopulation abdeckt. Das Konzept des log GSD zur Beschreibung der innerartlichen Variabilität ist im Bericht detailliert beschrieben. Bei den toxikokinetischen Daten war ein signifikanter Unterschied zwischen oralen und inhalativen Daten zu beobachten, wobei die Inhalationsdaten eine geringere Variabilität aufwiesen.
Die zu **toxikodynamischen** Unterschieden ausgewerteten Daten sind mit großen Unsicherheiten verbunden. Der Unterschied zwischen den niedrigsten Dosen oder

Konzentrationen, die Effekte zeigten, und den höchsten Dosen oder Konzentrationen ohne Effekte in anderen Individuen umschreibt einen Bereich von 3 bis 201.
Diese Ergebnisse wurden verglichen mit anderen publizierten Auswertungen. Substanzspezifische Daten zu toxikokinetischen Unterschieden und Fallbeispiele mit Modellierungen in physiologischen pharmakokinetischen (PBPK) Modellen weisen auf Extrapolationsfaktoren für die Toxikokinetik im Bereich von 1,5 bis 6 hin. Allerdings sind höhere Faktoren notwendig, um Unterschiede verursacht durch Polymorphismen, zum Beispiel bezüglich des Enzyms Cytochrom P450 2C9, zu berücksichtigen. Es ergab sich eine gute Übereinstimmung unserer Daten zur oralen Exposition mit den toxikokinetischen Daten der Datenbank von Hattis et al. (2002).
Abdo et al. (2015) veröffentlichten Daten zur toxikodynamischen Variabilität in vitro. Sie verwendeten in vitro-Toxizitätsdaten von permanenten Lymphoblastoidzellen humanen Ursprungs, die von mehr als 1000 Individuen von 5 Kontinenten stammten. Die Variabilität bezüglich der toxischen Wirkung von 179 Chemikalien innerhalb der Zelllinien konnte verwendet werden, um eine Verteilung zu toxikodynamischen Unterschieden zu etablieren. Diese zeigte eine gute Übereinstimmung mit Auswertungen von in vivo-Daten der Datenbank von Hattis et al. (2002).
Auf Grundlage unserer neu generierten Daten zur Toxikokinetik und der Daten von Abdo et al. (2015) war es möglich, datenbasierte Verteilungen zu toxikokinetischen und -dynamischen Unterschieden in der Empfindlichkeit in der erwachsenen Bevölkerung zu etablieren.

Der letzte Teilbericht des Projektes (**Synthesis report**) baute auf den Ergebnissen der vorangegangenen Phasen auf und analysierte die Schutzniveaus, die von den verschiedenen Methodiken zur Ableitung von Arbeitsplatzgrenzwerten und analogen Werten erreicht werden.
Die auf Basis einer großen aktuellen Datenbasis erstellten empirischen Verteilungen für Zeit-, Inter- und Intraspeziesextrapolation wurden benutzt, um parametrische Verteilungen abzuleiten, mit denen die Unsicherheit der Arbeitsplatzgrenzwerte beschrieben werden kann. Für die Daten zur Zeit-, Inter- sowie Intraspeziesextrapolation bezüglich toxikodynamischer (TD) Unterschiede erwiesen sich Lognormalverteilungen als geeignete Anpassungen an die empirischen Daten. Für interindividuelle toxikokinetische (TK) Unterschiede wurde die Verteilung der log GSD-Werte verwendet, um eine Verteilung abzuleiten.

$$\text{Verteilung der Verhältniswerte}_{\text{Intra,TK i}} = 10^{\log_{10} \text{GSD} * z_{1-i}}$$

wobei z_{1-i} der z-Score der Normalverteilung ist und den Anteil der Zielpopulation angibt, der durch die Extrapolation berücksichtigt werden soll (die Berechnungen wurden für einen Anteil von 95% und 99% der Zielpopulation durchgeführt, entsprechend einer verbleibenden Inzidenz von 5% bzw. 1%, siehe nachfolgende Tabelle).

Parametrische Verteilungen für einzelne Extrapolationsschritte, die in diesem Projekt auf Basis verschiedener Datenquellen abgeleitet wurden

Extrapolation	μ (Lognormal-verteilung)	σ (Lognormal-verteilung)	Median	75% Perzentil	95% Perzentil
Zeit: subakut/chronisch	1,31	1,05	3,71	7,52	20,85
Zeit: subchron. /chronisch	1,04	0,99	2,83	5,53	14,49
Interspezies	0,02	0,75	1,02	1,69	3.49
Intraspezies (TK und TD kombiniert) bei 1% Inzidenz		-	7,25	12,53	34,26
Intraspezies (TK und TD kombiniert) bei 5% Inzidenz		-	3,56	5,15	10,37

Diese Ergebnisse lassen sich z. B. für die Zeitextrapolation so interpretieren, dass für die Extrapolation von einer subakuten zu einer chronischen Studie für 50% der neu zu bewertenden Substanzen ein Faktor von 3,71 ausreichend ist. Bezüglich der Intraspeziesextrapolation beschreibt die Verteilung, dass für 50% der Substanzen (Median) zwischen der äquipotenten Dosis für das 50. und 5. Perzentil (d.h. den empfindlichsten 5%) der Population ein Faktor 3,56 liegt. Für den Unterschied zwischen dem 50. und dem 1. Perzentil (d.h. dem empfindlichsten 1% der Population) ergibt sich im Median ein Faktor von 7,25. Durch Vergleich der derzeit verwendeten Standardwerte für Extrapolationsfaktoren (dokumentiert u.a. in Report 1, Tabelle 2-5) mit diesen Verteilungen war es möglich, die Wahrscheinlichkeiten zu ermitteln, mit denen die Faktoren ihr Schutzziel erreichen. Es wurden bezüglich dieser Wahrscheinlichkeiten große Unterschiede beobachtet: hohe Wahrscheinlichkeiten ergaben sich für die Zeitextrapolation subakut zu chronisch und für die Interspeziesextrapolation, geringe Wahrscheinlichkeiten für die Zeitextrapolation subchronisch zu chronisch und für die Intraspeziesextrapolation (Report 10, Tabellen 3-1 bis 3-4). Große Unterschiede in den erreichten Wahrscheinlichkeiten zeigte auch der Vergleich der verschiedenen Ableitungsmethodiken: Mittels Monte-Carlo Simulation wurden die im Projekt erarbeitenden Verteilungen kombiniert. Die resultierende Gesamtverteilung wurde genutzt, um die Gesamtextrapolationsfaktoren der jeweiligen Systeme einzuordnen und die jeweilige Wahrscheinlichkeit der Erreichung des Schutzziels zu beschreiben (siehe z.B. Report 10, Abb. 3-12, für den Fall einer subakuten Studie).

Die nachfolgenden Tabellen dokumentieren die errechneten Wahrscheinlichkeiten, ausgehend von einer subakuten, subchronischen oder chronischen Studie.

Wahrscheinlichkeit, mit der die Standardextrapolationsfaktoren verschiedener Bewertungssysteme das Schutzziel erreichen (errechnet mittels probabilistischer Modellierung mit den in diesem Projekt erhaltenen Verteilungen) (POD aus einer subakuten Studie)*

OEL-Methodik	Effektart, Expositionspfad	Extrapolationsfaktoren (Zeit, Interspezies, Scaling, Intraspezies)	Vorgeschlagener Gesamtfaktor	Inzidenz (für Intraspeziesetrapolation)	Wahrscheinlichkeit
REACH/ RAC	systemisch, oral	6; 2,5; 4; 5	300	1 %	73,3 %
				5 %	88,0 %
	systemisch, Inhalation	6; 2,5; -; 5	75	1 %	73,3 %
				5 %	88,0 %
	lokal, Inhalation	6; 2,5, -; 5	75	1 %	73,3 %
				5 %	88,0 %
AGS	systemisch, oral	6; 5; 4; -	120	1 %	51,0 %
				5 %	70,3 %
	systemisch, Inhalation	6; 5; -; -	30	1 %	51,0 %
				5 %	70,3 %
	lokal, Inhalation	6; 5; -; -	30	1 %	51,0 %
				5 %	70,3 %
MAK*	systemisch, oral	6; 2; 4; -	48	1 %	28,0%
				5 %	45,3 %
	systemisch, Inhalation	6; 2; -; -	12	1 %	28,0%
				5 %	45,3 %
	lokal, Inhalation	6; 2; -; -	12	1 %	28,0%
				5 %	45,3 %
EU BPR	systemisch, oral	6; 10; -; 10	600	1 %	85,6 %
				5 %	95,0 %
	systemisch, Inhalation**				
	lokal, Inhalation	6; 2,5; -; 10	150	1 %	85,6 %
				5 %	95,0 %
ECETOC	systemisch, oral	6; 1; 4; 3	72	1 %	37,7 %
				5 %	56,8 %
	systemisch, Inhalation	6; 1; -; 3	18	1 %	37,7 %
				5 %	56,8 %
	lokal, Inhalation	1; 1; -; 3	3	1 %	6,5 %
				5 %	13,3 %

* neu gegenüber Report 10: Berechnungen für MAK hinzugefügt mit kombiniertem Intra- und Interspeziesfaktor 2

** inhalative POD werden in systemische Dosen konvertiert; es werden die gleichen AF wie für systemisch, oral verwendet

Wahrscheinlichkeit, mit der die Standardextrapolationsfaktoren verschiedener Bewertungssysteme das Schutzziel erreichen (errechnet mittels probabilistischer Modellierung mit den in diesem Projekt erhaltenen Verteilungen) (POD aus einer subchronischen Studie)*

OEL-Methodik	Effektart, Expositionspfad	Extrapolationsfaktoren (Zeit, Interspezies, Scaling, Intraspezies)	Vorgeschlagener Gesamtfaktor	Inzidenz (für Intraspeziesetrapolation)	Wahrscheinlichkeit
REACH/ RAC	systemisch, oral	2; 2,5; 4; 5	100	1 %	53,3 %
				5 %	73,0 %
	systemisch, Inhalation	2; 2,5; -; 5	25	1 %	53,3 %
				5 %	73,0 %
	lokal, Inhalation	2; 2,5; -; 5	25	1 %	53,3 %
				5 %	73,0 %
AGS	systemisch, oral	2; 5; 4; -	40	1 %	29,5 %
				5 %	47,7 %
	systemisch, Inhalation	2; 5; -; -	10	1 %	29,5 %
				5 %	47,7 %
	lokal, Inhalation	2; 5; -; -	10	1 %	29,5 %
				5 %	47,7 %
MAK*	systemisch, oral	2; 2; 4; -	16	1 %	12,1 %
				5 %	23,1 %
	systemisch, Inhalation	2; 2; -; -	4	1 %	12,1 %
				5 %	23,1 %
	lokal, Inhalation	2; 2; -; -	4	1 %	12,1 %
				5 %	23,1 %
EU BPR	systemisch, oral	2; 10; -; 10	200	1 %	70,8 %
				5 %	86,7 %
	systemisch, Inhalation**				
	lokal, Inhalation	2; 2,5; -; 10	50	1 %	70,8 %
				5 %	86,7 %
ECETOC	systemisch, oral	2; 1; 4; 3	24	1 %	18,7 %
				5 %	33,2 %
	systemisch, Inhalation	2; 1; -; 3	6	1 %	18,7 %
				5 %	33,2 %
	lokal, Inhalation	1; 1; -; 3	3	1 %	8,5 %
				5 %	17,1 %

* neu gegenüber Report 10: Berechnungen für MAK hinzugefügt mit kombiniertem Intra- und Interspeziesfaktor 2

** inhalative POD werden in systemische Dosen konvertiert; es werden die gleichen AF wie für systemisch, oral verwendet

Wahrscheinlichkeit, mit der die Standardextrapolationsfaktoren verschiedener Bewertungssysteme das Schutzziel erreichen (errechnet mittels probabilistischer Modellierung mit den in diesem Projekt erhaltenen Verteilungen) (POD aus einer chronischen Studie)*

OEL-Methodik	Effektart, Expositions-pfad	Extrapolations-faktoren (Zeit, Interspezies, Scaling, Intra-spezies)	Vorgeschla-gener Gesamt-faktor	Inzidenz (für Intraspezies-etrapolation)	Wahrschein-lichkeit
REACH/ RAC	systemisch, oral	-; 2,5; 4; 5	50	1 %	67,2 %
				5 %	89,5 %
	systemisch, Inhalation	-; 2,5; -; 5	12.5	1 %	67,2 %
				5 %	89,5 %
	lokal, Inhalation	-; 2,5; -; 5	12.5	1 %	67,2 %
				5 %	89,5 %
AGS	systemisch, oral	-; 5; 4; -	20	1 %	34,6 %
				5 %	62,0 %
	systemisch, Inhalation	-; 5; -; -	5	1 %	34,6 %
				5 %	62,0 %
	lokal, Inhalation	-; 5; -; -	5	1 %	34,6 %
				5 %	62,0 %
MAK*	systemisch, oral	-; 2; 4; -	8	1 %	10,1 %
				5 %	24,3 %
	systemisch, Inhalation	-; 2; -; -	2	1 %	10,1 %
				5 %	24,3 %
	lokal, Inhalation	-; 2; -; -	2	1 %	10,1 %
				5 %	24,3 %
EU BPR	systemisch, oral	-; 10; -; 10	100	1 %	85,2 %
				5 %	97,2 %
	systemisch, Inhalation**				
	lokal, Inhalation	-; 2,5; -; 10	25	1 %	85,2 %
				5 %	97,2 %
ECETOC	systemisch, oral	-; 1; 4; 3	12	1 %	18,8 %
				5 %	40,2 %
	systemisch, Inhalation	-; 1; -; 3	3	1 %	18,8 %
				5 %	40,2 %
	lokal, Inhalation	-; 1; -; 3	3	1 %	18,8 %
				5 %	40,2 %

* neu gegenüber Report 10: Berechnungen für MAK hinzugefügt mit kombiniertem Intra- und Interspeziesfaktor 2

** inhalative PODs are converted to systemic dose descriptors, AF as above for oral apply

Bei Betrachtung der Kombination aller Extrapolationsschritte ergab sich folgende Reihenfolge (mit abnehmender Wahrscheinlichkeit der Erreichung des Schutzziels):

BPR > RAC/REACH > AGS > MAK ≈ ECETOC.

(BPR: Biozidprodukte-Verordnung; RAC: Ausschuss für Risikobewertung (verantwortlich für die Ableitung von Vorschlägen zu Arbeitsplatzgrenzwerten in der

EU); REACH: ECHA-Leitlinien zur Ableitung von DNELs unter REACH; AGS: Ausschuss für Gefahrstoffe; MAK: MAK-Kommission; ECETOC: European Centre for Ecotoxicology and Toxicology of Chemicals).

Für die Methodik zur Bewertung von Pflanzenschutzprodukten kann ein ähnliches Schutzniveau angenommen werden wie im Falle der Biozidprodukte, obwohl in der Methodik nicht für alle Extrapolationsschritte Standardwerte angegeben werden.

Andersherum können mit der kombinierten Verteilung auch die Werte für den Gesamtfaktor bestimmt werden, die für die Erreichung eines bestimmten Schutzziels mit einer bestimmten Wahrscheinlichkeit stehen. Nachfolgend wird wiedergegeben, welche Gesamtextrapolationsfaktoren benötigt werden, um bei einer Effektinzidenz von 1% oder 5% in der Zielpopulation eine Wahrscheinlichkeit von 50%, 75% oder 95% zu erreichen.

Gesamtextrapolationsfaktoren, die zum Erreichen einer bestimmten Wahrscheinlichkeit (50% 75% 95%) benötigt werden; Szenarien 1% oder 5% verbleibende Inzidenz; POD aus subakuter, subchronischer oder chronischer Studie

Studiendauer	Inzidenz	Wahrscheinlichkeit	benötigter Gesamt-Extrapolationsfaktor
Subakut	1%	50%	28,9
		75 %	81,5
		95 %	389
	5%	50%	14,1
		75 %	36,6
		95 %	149
Subchronisch	1%	50%	22,1
		75 %	60,6
		95 %	280
	5%	50%	10,8
		75 %	27,2
		95 %	106
Chronisch	1%	50%	7,6
		75 %	16,2
		95 %	54,1
	5%	50%	3,8
		75 %	7,1
		95 %	18,8

Die parametrischen Verteilungen wurden auch mit Verteilungen verglichen, die in anderen probabilistischen Modellen vorgeschlagen wurden. Weiter wurde der Einfluss der Wahl des POD (NOAEL/LOAEL oder BMD) anhand von zwei Beispielen diskutiert, die probabilistisch mittels Monte-Carlo-Simulation bewertet wurden. Es wurde festgestellt, dass im Einzelfall das Verhältnis des NOAEL oder LOAEL zum BMD stark variieren kann. Da bei der Verwendung von NOAEL oder LOAEL als POD deren Unsicherheit nicht berücksichtigt wird, ist das Benchmark-Verfahren der bevorzugte Weg, um einen POD abzuleiten. Dazu ist die Festlegung einer BMR notwendig. Damit

wird auch der resultierende Arbeitsplatzgrenzwert klarer definiert (in Bezug auf die zugrundeliegenden kontinuierlichen oder quantalen Effekte und deren Effekthöhe beim POD). Die probabilistische Modellierung erlaubt es, die gesamte, die Unsicherheit beschreibende Verteilung der BMD in der Bewertung zu berücksichtigen.

Aus der Analyse resultierten mehrere Empfehlungen für die Ableitung von Arbeitsplatzgrenzwerten aus toxikologischen Daten:

Empfehlung 1:
Alle Methodiken sollten eindeutig die angestrebten Schutzziele benennen:
- der Anteil der Zielpopulation, der durch den Arbeitsplatzgrenzwert geschützt werden soll sowie
- die Wahrscheinlichkeit, mit der der Schutz vor adversen Wirkungen (wie durch den POD definiert) durch den Arbeitsplatzgrenzwert gewährleistet sein soll.

Empfehlung 2:
Das Benchmark-Verfahren sollte standardmäßig benutzt werden, um einen POD abzuleiten.

Empfehlung 3:
Die probabilistische Modellierung sollte weiterentwickelt und zum Vergleich und Adjustierung von deterministischen Verfahren verwendet werden.

Empfehlung 4:
Eine Ausweitung der Datenmenge zur interindividuellen Variabilität des Menschen nach Inhalation würde die Möglichkeit eröffnen, pfadspezifische Verteilungen zur Intraspeziesextrapolation abzuleiten.

Summary

Occupational exposure limits (OELs) are important tools for controlling and managing exposures to hazardous substances at the workplace. Various bodies at national and international level set OELs. Within the European Union, the Scientific Committee on Occupational Exposure Limits (SCOEL) proposed OELs, until the Committee for Risk Assessment (RAC) at the European Chemicals Agency (ECHA) took over this task in 2019. At the national level in Germany, two committees are engaged in the derivation of health-based occupational exposure limits: the Committee on Hazardous Substances (Ausschuss für Gefahrstoffe, AGS) and the Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area (the MAK Commission) of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). However, health-based guidance values for the workplace are not only established in the context of occupational safety and health legislation, but also in other regulatory areas. Under the EU chemicals legislation REACH (EC 2006) Derived No Effect Levels (DNELs) are used as part of the chemical safety assessment in registration dossiers prepared by registrants, and Acceptable Exposure Levels (AELs) and Acceptable Operator Exposure Levels (AOELs) for active substances are derived under the EU Biocidal Products Regulation (BPR) (EU 2021) and the EU Plant Protection Products (PPP) Directive (EC 2021), respectively. These latter values will be referred to here as “OEL-analogue values”.

The **methodological approaches** of several existing systems for deriving OELs and OEL-analogue values in Germany and the European Union were analysed and discussed with the objectives to identify differences and to make recommendations for harmonisation. The following topics were mainly addressed: Definition and scope of the values; databases evaluated and conditions for data searches; the methodology used to derive values for systemic endpoints; the methodology used to derive values for local respiratory effects. Respective observations made for the various frameworks were compiled in tables to facilitate the comparison (Report 1, Tables 2.2 – 2-5).
This analysis concludes that quantitative differences can occur at each individual step of the derivation process, most prominently when determining the point of departure (POD) and with the application of assessment factors. Recommendations focus on areas, where harmonisation is required for a consistent derivation of values, for example:

- The use of the BMD method for determining the POD
- The use of allometric scaling
- The size of assessment factors (e.g., with regard to time and intraspecies extrapolation).
- Necessity and provision of specific assessment factors for specific endpoints, such as reproductive toxicity (comprising of endpoints such as male and female fertility and developmental toxicity), respiratory toxicity and sensory irritation.

Lack of documentation on methodologies used for deriving OELs is generally seen as the reason for lack of transparency and a hindrance for harmonisation. Therefore, complete and detailed documentation of methods used is desirable, especially in the following areas:

- Selection of suitable POD (e.g., from several key studies or in a weight-of-evidence approach)
- Modification of the POD to account for exposure conditions of workers
- Step-by-step application of assessment factors and conditions for deviating from defaults.

In a separate report a state-of-the-art discussion of the application of **benchmark dose (BMD) modelling** in risk assessment is provided. Technically it is done with software or online applications that fit flexible mathematical models to experimental data. A mathematical function is obtained that describes the dose-response relationship of the experimental data. The effect size considered is called benchmark response (BMR). The corresponding benchmark dose (BMD) is the dose with an additional effect as defined by the BMR compared to the control. Confidence intervals express the uncertainty due to sampling and/or measurement error. The benchmark dose lower bound (BMDL) describes the lower bound of the (one-sided) 95th confidence limit of the BMD, the BMDU the upper bound. Currently, both the BMD and the BMDL are used as POD for further assessments, but we recommended to use the BMDL as it considers the uncertainty pertaining to the BMD. Guidance documents for the application of the BMD approach in risk assessment are available from the European Food Safety Agency (EFSA), the US Environmental Protection Agency (US EPA) and from the World Health Organisation (Environmental Health Criteria Document 240, Chapter 5). Recent developments in the tools based on PROAST, developed by the Dutch National Institute for Public Health and the Environment (RIVM) and in BMDS (by the US EPA) allow to account for model uncertainty by applying model averaging. We recommend to use model averaging for both quantal and continuous data.
In the guidance documents for the derivation of OEL values the benchmark approach is described in most cases only as an alternative for the extrapolation from a Lowest observed effect level (LOAEL) to a No adverse effect level (NAEL) in case a No observed adverse effect level (NOAEL) is missing. Background information on the approach is often limited and guidance on the application is missing. This contradicts the description of this method as “state of the science” for determining a point of departure (POD) for risk assessment.

The application of **BMD modelling** was demonstrated with ten **example** substances in a separate report. Five substances with quantal data and five with continuous data were selected. They include datasets where only LOAELs could be determined as well as one dataset consisting of epidemiological data.

Probabilistic approaches and their usefulness for toxicological hazard assessment were the subject of a further report. With probabilistic approaches the input data (BMD as the point of departure, assessment factors) for setting health-based exposure limits are described by distributions, expressing associated uncertainty (in the POD and the assessment factors) and variability (in the human population). These distributions are combined using probabilistic methods (Monte Carlo simulation), resulting in a distribution of the exposure limit. This approach requires to decide on the critical effect size or benchmark response (BMR) (in order to determine the POD, if a benchmark dose is used), on the percentage of the target population to be covered by the value and on the probability of achieving the defined protection level.

The distribution of the exposure limit then allows describing uncertainty and variability of the output, to better characterise the protection level achieved and to estimate the size and likeliness of health effects at higher concentrations. Two ready-to-use tools (for non-statisticians) are available:

- The APROBA tool developed in the frame of the WHO/IPCS project on “Evaluating and Expressing Uncertainty in Hazard Characterization” (WHO 2014).
- The Monte Carlo tool developed by EFSA

In view of these new developments, use of probabilistic approaches for hazard assessment is simplified and their use for method development and discussion of (combination of) deterministic factors, for comparison with standard assessments using deterministic factors and for refined assessments of complex cases is encouraged.

A short report summarises the possibilities and limitations of **route-to-route extrapolation**, which can be applied for deriving OELs or OEL-analogue values in the absence of reliable inhalation studies. Specific criteria and conditions have been proposed for applying route-to-route extrapolation by the Interdepartmental Group on Health Risks from Chemicals (IGHRC) and Geraets et al. (2014). Further, the ECHA Guidance document on Information Requirements and Chemical Safety Assessment, R.8 provides practical support for performing the individual extrapolation steps, starting from a route-specific point of departure (ECHA 2012). There is a general agreement that route-to-route agreement is applicable only, if the expected critical effects are of systemic, not local nature and if no differences exist that make predictions for the other route unreliable, for example a severe first-pass effect.

A key part of this research project discussed the extrapolation steps (time, inter- and intraspecies extrapolation) applied in deriving occupational exposure limits and aimed at improving their empirical database.
For **time** and **interspecies extrapolation**, distributions of ratios of dose descriptors were derived from studies of different length or species, exploiting studies of the US National Toxicology Program (NTP) and REACH registration data. NTP study data was manually assessed: For each evaluated report, doses (or concentrations in case of inhalation studies) corresponding to the NOAEL and/or LOAEL were determined for each type of study (species, sex, study duration) and type of endpoint (bodyweight, local and systemic effects). In case of REACH data, these doses were determined based on structured data provided by the European Chemicals Agency (ECHA), which was extracted from the REACH IUCLID database. The REACH database contains considerably more repeated dose toxicity studies than the NTP reports, however strict selection criteria, based on the study metadata reported in IUCLID, were necessary to ensure sufficient quality of the data used.
Using the obtained NTP and REACH datasets, for each substance the available subacute/subchronic, subchronic/chronic and subacute/chronic study pairs were compared by calculating the ratios $dose_{shorter\ study}/dose_{longer\ study}$. The resulting ratio distributions were further stratified according to study parameters to evaluate possible influencing factors. This evaluation of both datasets led to the conclusion that no consistent differences with regard to **time extrapolation** for the variables

- Route of application (oral, inhalation)

- Sex
- Species
- Toxicity endpoint types after inhalation (local, systemic)
- Target organs
- Substance classes (exemplary examined for two groups of substances with NTP data)

are evident. Due to observations on reporting quality of REACH data and the restrictions of a largely automated study evaluation (necessary due to the initially high number of studies), we consider the REACH database less reliable than the manual evaluation of NTP data. This conclusion is supported by a larger GSD (geometric standard deviation) for REACH data compared to NTP data, pointing to higher variability in this dataset. Therefore, we conclude that the combined dataset of ratios from oral and inhalation NTP data is adequate for proposing distributions for time extrapolation. The data are derived from studies on 256 substances, from which close to 400 (subacute/subchronic and subacute/chronic each) or more than 1200 (subchronic/chronic) ratios were calculated. For local effects in the respiratory tract after inhalation, the same distributions as for systemic effects are proposed, a conclusion which is supported by publications pointing to similar or higher ratios for local compared to systemic effects.
Our analysis presents one of the few data evaluations covering both sub-steps (subacute/subchronic and subchronic/chronic) as well as the full span (subacute – chronic) of the most frequently used time extrapolation steps. The GMs (geometric means) of ratios obtained for subacute versus chronic (GM 4.11) and subacute versus subchronic (GM 1.60) NTP studies fit very well to other evaluations published in recent years. The ratios for subchronic versus chronic studies (GM 2.93) are at the upper end of the range reported in recent evaluations. However, multiplication of GMs or medians of the two sub-steps yield values in agreement with the subacute – chronic ratios, indicating consistency in the three datasets.

Extrapolating from experimental animal studies to humans (**interspecies extrapolation**) is a key step in deriving OELs or analogue values. Existing methods either use default factors (often split in a toxicokinetic and a toxicodynamic part) and/or apply allometric scaling rules. Allometric scaling relates physiological and kinetic parameters to body weight raised to a certain power (for scaling according to caloric demand (also called basal metabolic rate scaling), the allometric exponent is 0.75).
Both NTP studies and repeated dose studies from REACH registrations were also evaluated with regard to interspecies differences. Dose descriptors were extracted as described above for “Time extrapolation”. In agreement with the predictions of allometric scaling the evaluation of oral studies from both datasets show that the larger species appears to be more susceptible, if doses are expressed per kg body weight (geometric means of dose ratio rats/mice: NTP: 0.40; REACH: 0.66; value expected according to allometric scaling: 0.59). Inhalation studies with rats and mice (for which the most datasets are available) show a similar susceptibility (geometric means of dose ratio rats/mice: NTP: 0.96; REACH: 1.09; expected value: 1), when exposure concentrations are compared, which is again in agreement with allometric principles. A relevant variability around the mean values is observed for all datasets (geometric standard deviations (GSDs): NTP: 3.6-3.8; REACH: 3.0-3.9).

These results are compared with existing empirical evaluations of toxicity and toxicokinetic data. Although the associated uncertainty does not allow to determine whether an allometric exponent of 0.7 or 0.75 is more adequate, the existing evaluations support the application of allometric scaling factors. Caloric demand (also called metabolic rate) scaling is recommended here, as it is supported by the empirical data as well as by mechanistic considerations. Caloric demand scaling also leads to the conclusion that no correction is required when deriving an OEL from an inhalation concentration as point of departure.
Based on the existing data a two-step approach is recommended:

- Correction (normalisation) of doses by allometric scaling factors derived from caloric demand scaling (for oral data only; no correction required for inhalation concentrations).
- Consideration of remaining uncertainty due to substance-to-substance variability by a distribution derived from the empirical datasets.

Due to its higher quality the empirical dataset derived from NTP data is preferable over the REACH dataset.

For describing differences between species regarding deposition and clearance of aerosols in the respiratory tract specific modelling approaches exist. The "**Human Equivalent Concentration**" (HEC) approach is a procedure to extrapolate an aerosol exposure concentration from an experimental animal study to an equivalent human concentration for a chronic workplace inhalation exposure scenario. In a separate report we discussed HEC calculations and possibilities of modelling the fate of solid particles in the lower respiratory tract with the "Multiple Pathway Deposition Model" (MPPD). Considerable uncertainty was identified for the quantification of species differences in deposition and elimination. One of the major drivers of uncertainty is the lack of knowledge on the most adequate unit for describing deposited doses ("normalisation"), e.g., dose per alveolar lung surface area or dose per the volume of alveolar macrophages.
Considering the high uncertainty inherent to the HEC procedure we propose to use the same distribution to account for uncertainties in interspecies extrapolation in case of particulates assessed by the HEC procedure as for systemic effects.

Inter-individual differences in susceptibility (**intraspecies extrapolation**) to chemical substances is another key aspect when deriving health-based guidance values. Such variability may have its origin in differences in toxicokinetics (i.e., inter-individual variation in internal dose at the same external exposure) or differences in toxicodynamics (i.e., inter-individual variation in responses of the target tissue to the same internal exposure). Various conditions are known to influence susceptibility, among them age, sex, genetics (e.g., polymorphisms of xenobiotic metabolising enzymes), epigenetic differences, and impaired health. Accordingly, quantification of the inter-individual variability for risk assessment purposes remains a challenge. Currently, methodologies for deriving OEL or analogue exposure limits use poorly justified default values. As part of this project new datasets on interindividual variability were compiled and evaluated, based on data from published human studies:

- 78 studies (68 of which could be evaluated quantitatively) on differences in toxicokinetics
- 25 studies on differences in toxicodynamics.

Variability in **toxicokinetic data** were characterised by log GSD values (the standard deviation of the logarithmic data). The median of log GSD values of the whole dataset was 0.146, equivalent to a factor of approx. 1.7 between the median and the 95th percentile of the population. The 95th percentile of log GSD of 0.355 corresponds to a factor of 3.8 to cover 95% of the population (the concept of log GSD for describing variability is further explained in the report). A significant difference between data from oral and inhalation exposure was observed, with lower variability for inhalation data.

The data on **toxicodynamics** are associated with large uncertainties. For the difference between the lowest dose or concentration showing effects in some individuals and the highest dose or concentration showing no effects in others, a range from 3 to 201 was observed.
These results were compared and evaluated with existing evaluations in the literature. Substance-specific data on toxicokinetic differences, as well as case studies using physiologically based pharmacokinetic (PBPK) modelling, result in toxicokinetic extrapolation factors in the range of 1.5 to 6, but higher factors are required for substances metabolised via polymorphic enzymes such as cytochrome P450 2C9. A high agreement was seen between the Hattis database (Hattis et al. 2002) on toxicokinetic differences and our data on oral exposures.
Abdo et al. (2015) published a database on toxicodynamic variability, using high-throughput screening data of immortalised lymphoblastoid cells from over 1000 individuals representing different populations from five different continents. The variability in the toxic responses observed in vitro in these cell lines to 179 chemicals could be used to derive a distribution for toxicodynamics, which is largely in agreement with human in vivo data from the Hattis database (Hattis et al. 2002).
In conclusion, our new database on toxicokinetic variability and the in vitro dataset of Abdo et al. (2015) on toxicodynamic variability allowed to establish data-derived distributions for toxicokinetic and toxicodynamic differences in susceptibility to chemical substances in the human adult population.

This last report of the project (**synthesis report**) built on the results of the previous project parts and analysed protection levels of the existing methodologies to derive occupational exposure limits (OELs) in the light of these results.
Using the empirical distributions for time, inter- and intraspecies extrapolation of toxicological data, parametric distributions were derived, which can be used in probabilistic modelling to describe the uncertainties of OELs. For time extrapolation, interspecies extrapolation (to be used in addition to allometric scaling) and intraspecies extrapolation (regarding differences in toxicodynamics, TD)) lognormal distributions were found, which fitted the empirical data well. For interindividual toxicokinetic (TK) differences the distribution of log GSD (geometric standard deviation) describing the toxicokinetic variability in the adult human population was used to establish a distribution.

$$\text{Distribution ratios}_{\text{Intra,TK i}} = 10^{\log_{10}\text{GSD} * z_{1-i}}$$

where $z_{1\text{-}i}$ is the z-Score of the normal distribution corresponding to the fraction of the population to be covered (calculations were performed for inclusion of 95% or 99% of the population, see the following table).

Parametric distributions for extrapolation steps, derived in this project based on several datasets

Extrapolation step	μ (log-normal distribution)	σ (log-normal distribution)	Median	75% percentile	95% percentile
Time: subacute/ chronic	1.31	1.05	3.71	7.52	20.85
Time: subchronic/ chronic	1.04	0.99	2.83	5.53	14.49
Interspecies	0.02	0.75	1.02	1.69	3.49
Combined (TK and TD) intraspecies at 1% incidence		-	7.25	12.53	34.26
Combined (TK and TD) intraspecies at 5% incidence		-	3.56	5.15	10.37

For example, the median factor of 3.71 observed for subacute to chronic extrapolation means that this factor is expected to be sufficiently protective in 50%of newly evaluated substances. The distribution for intraspecies extrapolation implies that a factor of 3.56 is sufficient for 50% of the substances to account for the differences in the equipotent doses of humans with median (50%) susceptibility and the 5% with the highest susceptibility. For the difference between the median susceptible humans and the 1% with the highest susceptibility a factor of 7.25 is required (for 50% of substances).
Currently used assessment factors (see Report 1, Table 2-5) were analysed by comparing with these distributions regarding the coverage (probability that the factor provides sufficient protection) achieved. Large differences in coverage provided were observed between different types of assessment factors (large coverage observed for subacute to chronic time extrapolation and interspecies extrapolation, lower coverage for subchronic to chronic time extrapolation and intraspecies extrapolation (Report 10, Tables 3-1 to 3-4). Also, large differences are obvious between the different frameworks: the distributions obtained in our project were combined by Monte Carlo Simulation and the resulting distribution was used to compare the various frameworks and to calculate the probability for achieving the protection goals (see Report 10, Fig. 3-12, for a subacute study).
The following tables show the calculated probabilities when starting from a subacute, subchronic or chronic study.

Probability achieved (covered fraction of the uncertainty distribution according to probabilistic modelling using the parameters presented in this project) by the default AF proposed in relevant OEL frameworks (POD from subacute study)*

OEL framework	Type of effect, route of application	Proposed AF (time, inter, scaling, intra)	Total AF proposed	Incidence for Intraspecies factor	Probability
REACH/RAC	systemic, oral	6, 2.5, 4, 5	300	1 %	73.3 %
				5 %	88.0 %
	systemic, inhalation	6, 2.5, -, 5	75	1 %	73.3 %
				5 %	88.0 %
	local, inhalation	6, 2.5, -, 5	75	1 %	73.3 %
				5 %	88.0 %
AGS	systemic, oral	6, 5, 4, -	120	1 %	51.0 %
				5 %	70.3 %
	systemic, inhalation	6, 5, -, -	30	1 %	51.0 %
				5 %	70.3 %
	local, inhalation	6, 5, -, -	30	1 %	51.0 %
				5 %	70.3 %
MAK*	systemic, oral	6, 2, 4, -	48	1 %	28.0%
				5 %	45.3 %
	systemic, inhalation	6, 2, -, -	12	1 %	28.0%
				5 %	45.3 %
	local, inhalation	6, 2, -, -	12	1 %	28.0%
				5 %	45.3 %
EU BPR	systemic, oral	6, 10, -, 10	600	1 %	85.6 %
				5 %	95.0 %
	systemic, inhalation**				
	local, inhalation	6, 2.5, -, 10	150	1 %	85.6 %
				5 %	95.0 %
ECETOC	systemic, oral	6, 1, 4, 3	72	1 %	37.7 %
				5 %	56.8 %
	systemic, inhalation	6, 1, -, 3	18	1 %	37.7 %
				5 %	56.8 %
	local, inhalation	1, 1, -, 3	3	1 %	6.5 %
				5 %	13.3 %

* new compared to Report 10: calculation for MAK added with combined inter- and intra AF of 2

** inhalative PODs are converted to systemic dose descriptors, same AF as for systemic, oral apply

Probability achieved (covered fraction of the uncertainty distribution according to probabilistic modelling using the parameters presented in this project) by the default AF proposed in relevant OEL frameworks (POD from subchronic study)*

OEL framework	Type of effect, route of application	Proposed AF (time, inter, scaling, intra)	Total AF proposed	Incidence for Intraspecies factor	Probability
REACH/RAC	systemic, oral	2, 2.5, 4, 5	100	1 %	53.3 %
				5 %	73.0 %
	systemic, inhalation	2, 2.5, -, 5	25	1 %	53.3 %
				5 %	73.0 %
	local, inhalation	2, 2.5, -, 5	25	1 %	53.3 %
				5 %	73.0 %
AGS	systemic, oral	2, 5, 4, -	40	1 %	29.5 %
				5 %	47.7 %
	systemic, inhalation	2, 5, -, -	10	1 %	29.5 %
				5 %	47.7 %
	local, inhalation	2, 5, -, -	10	1 %	29.5 %
				5 %	47.7 %
MAK*	systemic, oral	2, 2, 4, -	16	1 %	12.1 %
				5 %	23.1 %
	systemic, inhalation	2, 2, -, -	4	1 %	12.1 %
				5 %	23.1 %
	local, inhalation	2, 2, -, -	4	1 %	12.1 %
				5 %	23.1 %
EU BPR	systemic, oral	2, 10, -, 10	200	1 %	70.8 %
				5 %	86.7 %
	systemic, inhalation**				
	local, inhalation	2, 2.5, -, 10	50	1 %	70.8 %
				5 %	86.7 %
ECETOC	systemic, oral	2, 1, 4, 3	24	1 %	18.7 %
				5 %	33.2 %
	systemic, inhalation	2, 1, -, 3	6	1 %	18.7 %
				5 %	33.2 %
	local, inhalation	1, 1, -, 3	3	1 %	8.5 %
				5 %	17.1 %

* new compared to Report 10: calculation for MAK added with combined inter- and intra AF of 2

** inhalative PODs are converted to systemic dose descriptors, same AF as for systemic, oral apply

Probability achieved (covered fraction of the uncertainty distribution according to probabilistic modelling using the parameters presented in this project) by the default AF proposed in relevant OEL frameworks (POD from chronic study)*

OEL framework	Type of effect, route of application	Proposed AF (time, inter, scaling, intra)	Total AF proposed	Incidence for Intraspecies factor	Probability
REACH/RAC	systemic, oral	-, 2.5, 4, 5	50	1 %	67.2 %
				5 %	89.5 %
	systemic, inhalation	-, 2.5, -, 5	12.5	1 %	67.2 %
				5 %	89.5 %
	local, inhalation	-, 2.5, -, 5	12.5	1 %	67.2 %
				5 %	89.5 %
AGS	systemic, oral	-, 5, 4, -	20	1 %	34.6 %
				5 %	62.0 %
	systemic, inhalation	-, 5, -, -	5	1 %	34.6 %
				5 %	62.0 %
	local, inhalation	-, 5, -, -	5	1 %	34.6 %
				5 %	62.0 %
MAK*	systemic, oral	-; 2; 4; -	8	1 %	10.1 %
				5 %	24.3 %
	systemic, inhalation	-; 2; -; -	2	1 %	10.1 %
				5 %	24.3 %
	local, inhalation	-; 2; -; -	2	1 %	10.1 %
				5 %	24,3 %
EU BPR	systemic, oral	-, 10, -, 10	100	1 %	85.2 %
				5 %	97.2 %
	systemic, inhalation**				
	local, inhalation	-, 2.5, -, 10	25	1 %	85.2 %
				5 %	97.2 %
ECETOC	systemic, oral	-, 1, 4, 3	12	1 %	18.8 %
				5 %	40.2 %
	systemic, inhalation	-, 1, -, 3	3	1 %	18.8 %
				5 %	40.2 %
	local, inhalation	-, 1, -, 3	3	1 %	18.8 %
				5 %	40.2 %

* new compared to Report 10: calculation for MAK added with combined inter- and intra AF of 2
** inhalative PODs are converted to systemic dose descriptors, AF as above for oral apply

When the full set of assessment factors is compared with distributions combined by Monte Carlo simulation, the following sequence (with decreasing coverage) is observed:

BPR > RAC/REACH > AGS > MAK ≈ ECETOC.

(BPR: Biocidal Products Regulation; RAC: Committee for Risk Assessment (in charge of deriving OELs at the EU level); REACH: ECHA guidance for deriving DNELs under REACH; AGS: Ausschuss für Gefahrstoffe, German OEL system; MAK: MAK Commission (Permanent Senate Commission for the Investigation of Health Hazards

of Chemical Compounds in the Work Area of the Deutsche Forschungsgemeinschaft); ECETOC: European Centre for Ecotoxicology and Toxicology of Chemicals).

The framework for assessing plant protection products is assumed to provide similar protection goals as BPR, although default values are not available for all extrapolation steps.

The distributions combined by Monte Carlo simulation can also be used to calculate the total AF required to achieve a certain protection goal. The following table shows which total AFs are required to achieve the protection goals to protect 95% (5% incidence) or 99% (1% incidence) of the target population with a probability of 50%, 75% or 95%.

Total extrapolation factors needed to achieve a defined probability (50%, 75% and 95%) for the scenarios 1% or 5% incidence and departing from a subacute, subchronic or chronic study

Duration of key study	Incidence	Probability	Total AF needed
subacute	1%	50%	28.9
		75 %	81.5
		95 %	389
	5%	50%	14.1
		75 %	36.6
		95 %	149
subchronic	1%	50%	22.1
		75 %	60.6
		95 %	280
	5%	50%	10.8
		75 %	27.2
		95 %	106
chronic	1%	50%	7.6
		75 %	16.2
		95 %	54.1
	5%	50%	3.8
		75 %	7.1
		95 %	18.8

The parametric distributions obtained were also compared with those proposed for use in other probabilistic models. Further, the influence of the point of departure (POD: NOAEL/LOAEL or BMD) was discussed and exemplified with two example substances modelled probabilistically by Monte Carlo simulation. The position of the NOAEL or LOAEL relative to the BMD can vary substantially. As the uncertainty inherent to the NOAEL or LOAEL is not considered when using these PODs, benchmark dose modelling is the preferred way to derive the POD. It requires defining a benchmark response and, hence, also allows a clearer definition of the OEL (regarding the type of - quantal or continuous - effect and the effect size at the POD). Probabilistic modelling allows to use the full distribution of the BMD for the assessment.

Several recommendations for deriving OELs or analogue values were derived from the findings:

Recommendation 1:
All OEL derivation frameworks should clearly define their protection goals by stating:
- The fraction of the exposed population covered by the OEL
- The probability with which they intend to provide protection from adverse effects (as defined by the POD)

Recommendation 2:
Benchmark dose modelling should be used as the default procedure to derive a POD

Recommendation 3:
Probabilistic models should be further developed and used for benchmarking against deterministic methodologies to test them

Recommendation 4:
Increasing and improving the database on inter-individual variability in human inhalation studies might allow to establish route-specific distributions for intraspecies variability.

Literatur

Abdo, N., Xia, M., Brown, C.C., Kosyk, O., Huang, R., Sakamuru, S., Zhou, Y.H., Jack, J.R., Gallins, P., Xia, K., Li, Y., Chiu, W.A., Motsinger-Reif, A.A., Austin, C.P., Tice, R.R., Rusyn, I., Wright, F.A. (2015). Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 genomes high-throughput screening study. Environmental Health Perspectives 123, 458-466.

Bhat, V.S., Meek, M.E., Valcke, M., English, C., Boobis, A., Brown, R. (2017). Evolution of chemical-specific adjustment factors (CSAF) based on recent international experience; increasing utility and facilitating regulatory acceptance. Critical Reviews in Toxicology 47, 729-749.

Dankovic, D.A., Naumann, B.D., Maier, A., Dourson, M.L., Levy, L.S. (2015). The scientific basis of uncertainty factors used in setting occupational exposure limits. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S55-S68.

Deveau, M., Chen, C.P., Johanson, G., Krewski, D., Maier, A., Niven, K.J., Ripple, S., Schulte, P.A., Silk, J., Urbanus, J.H., Zalk, D.M., Niemeier, R.W. (2015). The Global Landscape of Occupational Exposure Limits--Implementation of Harmonization Principles to Guide Limit Selection. Journal of Occupational and Environmental Hygiene 12 Suppl 1, S127-144.

EC, European Commission. (2021). Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Consolidated text. 02009R1107 — EN — 27.03.2021 — 007.002, http://data.europa.eu/eli/reg/2009/1107/2021-03-27.

EC, European Community. (2006). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC, OJ L 396, 30.12.2006.

ECHA, European Chemicals Agency. (2012). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1, November 2012, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

EU, European Union. (2021). REGULATION (EU) No 528/2012 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 May 2012 concerning the making available on the market and use of biocidal products. 02012R0528 — EN — 10.06.2021 — 005.001, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02012R0528-20210610.

Falk-Filipsson, A., Hanberg, A., Victorin, K., Warholm, M., Wallén, M. (2007). Assessment factors - applications in health risk assessment of chemicals. Environmental Research 104, 108-127.

Geraets, L., Bessems, J.G., Zeilmaker, M.J., Bos, P.M. (2014). Human risk assessment of dermal and inhalation exposures to chemicals assessed by route-to-route extrapolation: The necessity of kinetic data. Regulatory Toxicology and Pharmacology 70, 54-64.

Hattis, D., Baird, S., Goble, R. (2002). A Straw Man Proposal for a Quantitative Definition of the RfD, in Final Technical Report, U.S. Environmental Protection Agency STAR grant # R825360, Human Variability in Parameters Potentially Related to Susceptibility for Noncancer Risks. Full version available at http://www2.clarku.edu/faculty/dhattis ; shortened version in Drug and Chemical Toxicology 2002; 25:403-436.

Kalberlah, F., Heine, K. (2015). DNELs und Arbeitsplatzgrenzwerte - ein verlässlicher Maßstab zum Schutz vor beruflichen Gesundheitsgefährdungen? Kap. 9/6.5. In: U. Welzbacher (Ed), Betriebliches Gefahrstoffmanagement, p. 36 S.

Kalberlah, F., Schneider, K. (1998). Quantification of Extrapolation Factors. Final Report of the Research Project No. 116 06 113 of the Federal Environmental Agency, Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund Fb 797 Wirtschaftsverlag NW Bremerhaven.

Lehman, A.J., Fitzhugh, O.G. (1954). 100-fold margin of safety. Quarterly Bulletin - Association of Food & Drug Officials of the United States 18, 33-35.

Maier, A., Lentz, T.J., MacMahon, K.L., McKernan, L.T., Whittaker, C., Schulte, P.A. (2015). State-of-the-science: the evolution of occupational exposure limit derivation and application. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S4-S6.

Nies, E., Musanke, U., Püringer, J., Rühl, R., Arnone, M. (2013). DNELs for workplaces - observations from an inspection of the DGUV DNEL list. Gefahrstoffe - Reinhaltung der Luft 73, 455-462.

Schenk, L. (2010). Comparison of data used for setting occupational exposure limits. International Journal of Occupational and Environmental Health 16, 249-262.

Schenk, L., Hansson, S.O., Rudén, C., Gilek, M. (2008). Occupational exposure limits: A comparative study. Regulatory Toxicology and Pharmacology 50, 261-270.

Schenk, L., Johanson, G. (2011). A quantitative comparison of the safety margins in the European indicative occupational exposure limits and the derived no-effect levels for workers under REACH. Toxicological Sciences 121, 408-416.

Schenk, L., Johanson, G. (2018). Use on uncertainty factors by the European Commission Scientific Committee on Occupational Exposure Limits: a follow-up. Critical Reviews in Toxicology 48, 513-521.

Schenk, L., Johanson, G. (2019). Will worker DNELs derived under the European REACH regulation extend the landscape of occupational exposure guidance values? Archives of Toxicology 93, 1187-1200.

Schenk, L., Palmen, N., Theodori, D. (2014). Evaluation of worker inhalation DNELs. Part A: Quality assessment of a selection of DNELs. Part B: Discussion paper on the possibilities to improve the overall quality of DN(M)ELs. RIVM Letter report 110001001/2014, RIVM, National Institute for Public Health and the Environment.

US EPA, Environmental Protection Agency. (2014). Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. EPA/100/R-14/002F, U.S. Environmental Protection Agency, Risk Assessment Forum, Office of the Science Advisor, Washington, DC. https://www.epa.gov/sites/production/files/2015-01/documents/ddef-final.pdf.

WHO, World Health Organization. (1994). Environmental Health Criteria 170, Assessing Human Health Risks of Chemicals: Derivation of Guidance Values for Health-Based Exposure Limits, IPCS International Programme on Chemical Safety; World Health Organization Geneva.

WHO, World Health Organization. (2014). Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. IPCS Harmonization Project Document 11, WHO, Geneva, Switzerland. http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf.

REPORT 1:
Comparison of Methods for Deriving OELs

RESEARCH PROJECT F2437:
Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Summary

In this report the methodological approach of several existing systems for deriving OELs and OEL-analogue values in Germany and the European Union are analysed and discussed with the objective to identify differences and to make recommendations for harmonisation.

The report addresses methodological details with regard to

- Definition and scope of the values
- Databases evaluated and how data are searched
- The methodology used to derive values for systemic endpoints
- The methodology used to derive values for local respiratory effects
- Specific provisions for carcinogens with presumed thresholds.

Existing guidance documents and methodological descriptions are analysed and observations are documented in tabular form. These observations are discussed in a broader context, taking into account the international discussion on harmonising occupational exposure limits and similar values such as workers DNELs. This analysis concludes that quantitative differences can occur at each individual step of the derivation process:

- data searches and selection of databases for evaluation
- prioritising information (e.g., weighing human versus animal data) and selection of key studies
- determination of the POD(s)
- application of assessment factors (and deviation from defaults)
- adjustment to human (exposure) conditions
- weight-of-evidence considerations of additional information.

Recommendations given focus on areas, where harmonisation is required for a consistent derivation of values, for example:

- The use of the BMD method for determining the POD (and for using LOAELs as POD)
- The use of allometric scaling
- Quantification of assessment factors (e.g., with regard to time and intraspecies extrapolation).

- Necessity and provision of specific assessment factors for specific endpoints, such as reproductive toxicity (comprising of endpoints such as male and female fertility and developmental toxicity), respiratory toxicity and sensory irritation.

A persistent and unanimous topic in literature is the absence of (sufficient) guidance. Lack of documentation on methodologies used for deriving OELs is generally seen as the reason for lack of transparency and a hindrance for harmonisation. Therefore, complete and detailed documentation of methods used is desirable, which should not only describe the default approach, but also when and how deviation from defaults is possible. More precisely, the following areas are identified where more detailed guidance is required for guiding assessors, such as:

- Selection of suitable POD (e.g., from several key studies or in a weight-of-evidence approach)

- Modification of the POD to account for exposure conditions of workers

- Step-by-step application of assessment factors and conditions for deviating from defaults.

Generally spoken, in order to achieve the overall goal of transparency and harmonisation, the following elements are necessary: availability of detailed guidance, compliance with guidance and transparent documentation of evaluations.

Abbreviations

AEL	Acceptable Exposure Levels
AAEL	Acute Acceptable Exposure Levels
AF	Assessment factor
AGS	Ausschuss für Gefahrstoffe
AGW	Arbeitsplatzgrenzwert
AIC	Akaike information criterion
ANSES	Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail
AOEL	Acceptable Operator Exposure Levels
AAOEL	Acute Acceptable Operator Exposure Levels
APROBA	Approximate probabilistic analysis
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
BBMD	Bayesian Benchmark Dose
BMD	Benchmark dose
BMDL	Benchmark dose lower bound
BMDU	Benchmark dose upper bound
BMR	Benchmark response
BMDS	Benchmark dose software
BOELV	Binding occupational exposure limit values
BPR	Biocidal products regulation
BS	Bootstrapping
CDS	Cumulative distribution function
CES	Critical effect size
CSAF	Chemical-specific adjustment factors

DFG	Deutsche Forschungsgesellschaft
DMEL	Derived minimal effect level
DNEL	Derived no effect level
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
ED10	Effective dose 10% (dose corresponding to a 10% increase in an adverse effect, relative to the control response)
EFSA	European Food Safety Authority
GM	Geometric mean
GSD	Geometric standard deviation
GV	Guidance value
IPCS	WHO's International Programme on Chemical Safety
IRIS	Integrated Risk Information System
LOAEC	Lowest observed adverse effect concentration
LOAEL	Lowest observed adverse effect level
MAK	Maximale Arbeitsplatzkonzentration
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MCRA	Monte Carlo Risk Assessment
MPPD	Multiple path particle dosimetry (model)
NAEC	No adverse effect concentration
NAEL	No adverse effect level
NOAEC	No observed adverse effect concentration
NOAEL	No observed adverse effect level
OEL	Occupational exposure limit

PBPK	Physiology-based pharmacokinetic (model)
PDF	Probability density function
POD	Point of departure
PPP	Plant protection products
PROAST	Dose-response modelling software by RIVM
QSAR	Quantitative structure activity relationship
RAC	Committee for Risk Assessment
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals,
RfD	Reference dose
RIVM	Dutch National Institute for Public Health and the Environment
SC	EFSA's Scientific Committee
SCOEL	Scientific Committee on Occupational Exposure Limits
STEL	Short-term exposure limit
SD	Standard deviation
TD	Toxicodynamics
TK	Toxicokinetics
TRGS	Technische Regeln für Gefahrstoffe
US EPA	Environmental Protection Agency in the US
WHO	World Health Organisation

1 Introduction

The research project F2437 “Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels” was initiated with the overall objectives to analyse and update the scientific basis of setting OEL(-analogue) values and to contribute to harmonising existing methods by

- analysing and comparing existing methods
- making transparent the protection goals achieved by the assessments
- proposing suitable distributions for extrapolation and evaluation steps, based on up-to-date evaluation of data
- presenting and discussing new methodological approaches such as probabilistic methods.

In this part of the project a detailed comparison of existing methods to derive occupational exposure limits (OELs) or similar values in other regulatory areas is performed. In the scope of this comparison are the following type of regulatory values and their methodology:

- Official and legally binding German occupational exposure limits („Arbeitsplatzgrenzwerte“, AGW) (AGS 2010, 2018)

- Non legally binding OELs („maximale Arbeitsplatzkonzentrationen“, MAK values) derived by DFG-Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe („MAK-Kommission“) (DFG 2018)

- OELs as derived by the Scientific Committee on Occupational Exposure Limit Values (SCOEL) (SCOEL 2013, 2017)

- OELs as derived by the Committee on Risk Assessment (RAC) (ECHA 2019)

- Derived no effect levels (DNELs) for workplaces under REACH (ECHA 2012)

- Proposals for deriving DNELs by the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) (ECETOC 2003, 2010)

- Acceptable Exposure Levels (AEL values) for biocidal products according to the Biocidal Products Regulation (BPR) in the EU (ECHA 2017)

- Acceptable Operator Exposure Levels (AOEL values) for active substances in plant protection products according to the EU Plant Protection Products (PPP) Directive (EC 2006).

This comparison is mainly based on the written methodological documentation available but includes practical knowledge and scientific publications on the subject.

As research project F2437 focusses on threshold-based values, regulatory values for non-threshold substances such as genotoxic carcinogens are not subject to this discussion. But borderlines, e.g. how carcinogens with assumed thresholds are dealt with, are described.

In the frame of this report, the term OEL is used for values derived in the context of occupational safety and health regulations, i.e. values derived by the German committees, SCOEL and RAC (under assignments of DG Employment, Social Affairs and Inclusion). REACH DNELs, AOELs for pesticides and AELs for biocides derived for workers are summarised with the term “OEL-analogue values”. Note that SCOEL is no longer active. From 2019, the scientific evaluation of chemical substances at the workplace on behalf of the European Commission is performed by the Risk Assessment Committee (RAC) of the European Chemicals Agency (ECHA).

A strict comparison is not always possible, as differences exist due to historical developments and overall intentions. A general characteristic is that OELs are more workplace-situation driven, whereas other frameworks, e.g. REACH DNELs, are substance-driven. For example, OEL systems typically can include rules and values for inert dust exposure or for activity-related situations (e.g. welding), whereas under REACH DNELs for a large number of chemical substances were derived, often with limited consideration of specific use situations. Differences also exist in the level of detail and guidance given in the respective methodological documents. In the following table the guidance documentation available for the various OELs and OEL-analogue values are listed.

Table 1-1 Sources and background documentation consulted

System	Documentation
REACH Regulation	- ECHA Guidance on information requirements and chemical safety assessment R.8: Characterisation of dose [concentration] - response for human health. Version 2.1. November 2012 (ECHA 2012) - ECHA Guidance on information requirements and chemical safety assessment. Part B: Hazard Assessment. Version 2.1. December 2011 (ECHA 2011)
RAC OEL methodology	- ECHA Guidance on information requirements and chemical safety assessment R.8: Characterisation of dose [concentration] - response for human health. Version 1.0. Appendix R.8-17 Guidance for proposing Occupational Exposure Limits. August 2019 (ECHA 2019) - Joint Task Force ECHA Committee for Risk Assessment (RAC) and Scientific Committee on Occupational Exposure Limits (SCOEL) (2017a) on Scientific aspects and methodologies related to the exposure of chemicals at the workplace. Final version. 28 February 2017 (ECHA/RAC-SCOEL 2017a) - Joint Task Force ECHA Committee for Risk Assessment (RAC) and Scientific Committee on Occupational Exposure Limits (SCOEL) on Scientific aspects and methodologies related to the exposure of chemicals at the workplace. TASK 2. 6 December 2017. Final report (ECHA/RAC-SCOEL 2017b)
SCOEL	- EC, European Commission, Methodology for derivation of occupational exposure limits of chemical agents. The

	General Decision-Making Framework of the Scientific Committee on Occupational Exposure Limits (SCOEL) (2017)
AGS – German OELs	- AGS, Ausschuss für Gefahrstoffe (2010). Bekanntmachung zu Gefahrstoffen. Kriterien zur Ableitung von Arbeitsplatzgrenzwerten. BekGS 901 (AGS 2010) - Technische Regeln für Gefahrstoffe, Arbeitsplatzgrenzwerte (TRGS 900), 2006 (AGS 2018) - Ausschuss für Gefahrstoffe (2013). Leitfaden zur Quantifizierung stoffspezifischer Expositions-Risiko-Beziehungen und von Risikokonzentrationen bei Exposition gegenüber krebserzeugenden Gefahrstoffen am Arbeitsplatz, (Anlage 3 zu TRGS 910) (AGS 2013)
DFG MAK	- MAK-und BAT-Werte-Liste: Ständige Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe. Mitteilung 55 (DFG 2019)
ECETOC	- ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals, Technical Report No. 110. Guidance on Assessment Factors to Derive a DNEL, Brussels, Belgium (2010) - ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals, Technical Report No. 86. Derivation of Assessment Factors for Human Health Risk Assessment, Brussels, Belgium (2003)
Plant Protection Products Directive	- Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data, EFSA Scientific Committee, EFSA Journal 2012;10(3):2579 (EFSA 2012) - EFSA (European Food Safety Authority), 2006. Opinion of the Scientific Panel on plant protection products and their residues on the request from the Commission on the Guidance Document (GD) for the establishment of acceptable operator exposure levels (AOELs). The EFSA Journal (2006) 345, 1-12 (EFSA 2006) - European Commission, Health & Consumer Protection Directorate-General, Directorate E – Safety of the food chain, E3 - Chemicals, Contaminants, Pesticides. Draft Guidance for the setting and application of acceptable operator exposure levels (AOELs), SANCO 7531 - rev.10, 7 July 2006 (EC 2006)
EU Biocidal Products Regulation	- ECHA Guidance on the Biocidal Products Regulation. Volume III Human Health - Assessment & Evaluation. (Parts B+C). Version 4.0. December 2017 (ECHA 2017)

The following section 2.1 to 2.5 summarise the observations made when comparing the existing guidance documents (**Table 1-1**) step by step for the main subjects

- Definition and scope of values
- Instructions for data search and data evaluation
- Methodology for deriving values for systemic effects
- Methodology for deriving values for local effects in the respiratory tract
- Handling of carcinogens with assumed thresholds.

The conclusions drawn on these subjects are based on detailed information obtained from analysing the guidance documents as documented in the tables in Annex 1.

The observations made are discussed in a broader context including other sources in the respective subsections of section 3.

2 Comparison of methods

2.1 Definitions and scope of values

In this section it is assessed how OELs and OEL-analogue values are defined with regard to their scope, the type/level of protection they strive to achieve, and whether critical endpoints such as developmental toxicity and sensitive subgroups are included.

Although it is not explicitly mentioned in all guidance documents, a major difference between OELs and the other values is that the former are derived by expert committees, whereas the guidance documents for OEL-analogue values are meant to be used by (numerous) individual assessors (e.g. from companies preparing REACH registration dossiers) and authorities alike. This has implications also on the level of detail in guidance documents: for independent assessors detailed guidance is mandatory for achieving a common approach, whereas committees may develop a common (committee-internal) understanding and routine of deriving values even in absence of detailed guidance.

Also, it should be noted that comparing documented guidance is different to a comparison of the outcome of the assessment processes. It can be expected that assessments conducted by a large number of assessors, as it is the case with DNELs derived under REACH by individual consultants or companies, will show larger variability compared to assessments performed by groups (committees) of experts with limited fluctuation of memberships. In section 3.3 some publications are discussed, which include quantitative comparisons of derived values. These comparisons give some insight into the practice of using the guidance documents rather than comparing the documents themselves.

2.1.1 Observations

The following table summarises key observations made in the detailed analysis as documented in Annex 1.

Table 2-1 Scope and definition of OELs and OEL-analogue values

Subject	Observation
Target population	By their very nature OELs such as those derived by SCOEL or the German MAK-Kommission are specific for workers; in contrast, AOELs derived for active substances in plant protection products are supposed to be used to assess exposure of operators, but also bystanders and residents; also AELs derived for biocidal active substances are to be used for professional and non-professional users of the products. Target group-specific DNELs are derived under REACH
Specification of values and unit(s)	OELs derived by workplace committees or inhalative DNEL values for workers derived in the REACH context are expressed as air concentrations (units mg/m³ or ppm). In contrast, AOELs and AELs for pesticides and biocidal active substances are typically given as systemic doses in mg/kg body weight/day. As a consequence, values expressed as air concentrations relate to a specific exposure scenario (exposure assumed during 8 h per day, 5 days per week, 48 weeks per year,40 years), whereas AOELs and AELs are expressed as the internal (absorbed) dose, which results from exposure via all routes. SCOEL and the German MAK Commission use the so-called "preferred value approach", as the numerical values 1, 2 or 5, multiplied by powers of ten. In some cases, this may be a cause for slight numerical differences to others.
Protection goal	Similar definitions apply to the various values (although the wording differs): all values aim at protecting workers, including sensitive individuals or groups, from experiencing adverse health effects
Does include developmental toxicity?	Differences exist regarding the quantitative consideration of developmental toxic effects when deriving OEL(-analogue) values: This endpoint is explicitly to be considered for deriving DNELs, AOELs or AELs. Also the new guidance documents for deriving OELs by RAC and SCOEL ask for quantitative inclusion of developmental toxicity data; however this is a change in SCOEL's philosophy, as the 2013 guidance document (SCOEL 2013) explains that "OELs established to protect adults cannot a priori guarantee the absence of pre- or post-natal adverse effects". The German systems assign substances to pregnancy groups, which indicate whether the OEL provides sufficient protection also for the unborn child.
Does consider endocrine disruptive activity?	Specific attention to endocrine disrupting activities is given in the regulatory areas of REACH, plant protection products and biocides, although for the purpose of a qualitative assessment (i.e. identification of endocrine disruptors) rather than a quantitative consideration when deriving values. Endocrine effects are not specifically mentioned in guidance documents from workplace committees.
Does consider respiratory sensitisation and sensory irritation?	Local effects in the respiratory tract (including sensory irritation) as well as respiratory sensitisation are given more attention in the area of OELs derived by workplace committee; guidance documents in the

Subject	Observation
	plant protection and biocidal product area do not specifically mention these endpoints, whereas ECHA's REACH guidance asks for a qualitative assessment; OEL-deriving bodies assign notations for respiratory sensitisation, but in addition, on a case-by-case basis where data allow, dose-response data can also be used for setting the OEL (see Schenk and Johanson (2019))
STEL values derived?	Workplace committees typically consider peak exposures by deriving short-term exposure levels (STELs) (or respective exceeding factors) for 15 min exposure periods; in other regulatory areas short-term values covering varying (often longer) exposure periods can be derived, but short-term inhalation DNELs are derived on a case-by case basis only under REACH
Skin notations	Skin notations are the preferred way to control substances with high uptake via skin in OEL systems. Quantitative route-specific DNELs are derived under REACH. The AOELs or AELs derived for pesticides and biocidal active substances are systemic absorbed doses also used to assess dermal exposure

2.1.2 Conclusions

Although the principal objective (to provide protection from adverse effects to the target population, including sensitive individuals) is common to all values investigated, there are differences. The most obvious differences can be observed between workplace OELs on one side and the OEL-analogue values derived in other regulatory areas on the other side:

- AOEL and AEL values for pesticides and biocides, respectively, are also used to assess exposures of persons not exposed in a professional context. They are expressed as absorbed systemic doses, whereas in all other areas air concentrations are typically derived. OELs (but also DNELs for workers inhalation exposure under REACH) are air concentrations defined for conditions of a specific workplace scenario (8 hours per day, 5 days per week).

- Different endpoints are in the focus in the various regulatory areas. Due to regulatory requirements endocrine disrupting activities are addressed explicitly in the area of pesticides and biocides, and may be a topic also for REACH substances. In the area of pesticides and biocides, identification as an endocrine disrupting chemical in most cases will lead to non-acceptance of the substance; hence, respective evidence is used for hazard characterisation and not for deriving OEL(-analogue) values.

- In contrast, more attention is given to effects in the respiratory tract such as sensory irritation and respiratory sensitisation by committees deriving OELs; notations for respiratory sensitisation are assigned in OEL systems.

- Developmental toxicity is a key endpoint to be considered in the REACH area and for biocides and pesticides and is mentioned as such also in the new guidance documents for deriving OELs by RAC and SCOEL; however, in the case of SCOEL this is a new development, which might not be mirrored by the substance-specific evaluations: the 2013 guidance document (SCOEL 2013) does not define OELs to provide protection against developmental toxicity in all cases (see table above). The German systems assign substances to pregnancy groups, which indicate whether the OEL provides sufficient protection also for pregnant women.

- Also, the dermal route is dealt with differently: skin notations are assigned to control substances with high uptake via skin in OEL systems (however, on a case-by-case basis high skin penetration might also lead to setting lower OELs), whereas quantitative route-specific DNELs are derived under REACH; AOELs or AELs derived for pesticides and biocidal active substances are systemic absorbed doses also used to assess dermal exposure.

- Further differences are noticed with regard to dealing with short-term exposure situations: OEL systems typically provide STEL values (or corresponding exceedance factors) for 15 minute exposure periods, whereas under REACH short-term values for varying durations are derived on a case-by case basis; acute reference doses for pesticides and biocides are typically derived for the oral pathway only.

2.2 Data search and evaluation

Documents were checked for information on the methods and requirements for performing data searches, identifying key information and their evaluation and documentation.

ECETOC publications specifically aim at providing justifications for default factors to be used for DNEL derivation. Therefore, information on information searches and sources are absent in these documents (ECETOC 2003, 2010). This holds also true for the AGS documentation on default extrapolation factors (AGS 2010).

2.2.1 Observations

The following table summarises key observations made in the detailed analysis as documented in Annex 1.

Table 2-2 Observations on retrieval and evaluation of data

Subject	Observation
Requirements for information searches and type of information to be used	Guidance on how to carry out data searches and relevant sources varies substantially in the level of detail given; in its new guidance document SCOEL provides a detailed description on how to conduct searches; in the REACH and biocides area an own guidance document (R.3) exist, which covers this topic
Assessment of data quality	The ECHA Guidance on Information Requirements and Chemical Safety Assessment, R.4, provides principles (Klimisch scores) used in the REACH and biocides area, as well as by RAC for setting OELs; no specific criteria are given by others
Identification of critical effects / key studies and weight-of-evidence (WoE) considerations	Identification of the key study is a critical element in all systems and weight-of-evidence approaches are mentioned by most as part of the evaluation process; some documentation explicitly state that the lowest NOAEL might not necessarily lead to the lowest OEL and therefore all steps of the derivation should be considered for selecting the key study.
Application of read-across and QSAR	These tools are specifically addressed in the REACH and biocides area (and also by RAC for setting OELs)
Use of human data	Considered important in all systems
Severity of effects	Generally considered as part of the evaluation process; some recommend specific or higher assessment factors for severe effects (for LOAEC-NAEC extrapolation: MAK (for sensory irritation) and ECETOC; for reproductive toxicity: BPR Guidance)
Update of evaluations	The ECHA Guidance R.8 and SCOEL require to update evaluations when relevant new information becomes available; MAK Commission and the respective working group of AGS publish lists of substances in (re-)evaluation; no specific procedures are documented by others
Documentation requirements for data and assessments	All derivations are documented, but there are few requirements fixed on the content and level of detail; those derivations performed as regulatory requirements (REACH, biocides) follow the rules for the documents to be submitted (to be noted: REACH DNEL derivations are not publicly available in every case)

2.2.2 Conclusions

- With regard to data search large differences exist in the guidance provided: the ECHA Guidance document on Information Requirements and Chemical Safety Assessment R.3 is applicable for REACH and BPR dossiers and SCOEL in its new guidance document also provides detailed recommendations; for plant protection products a large set of experimental data needs to be provided and the assessment relies to a larger part on the submitted data; searches for publicly available data might be less relevant (nevertheless, for reapplications external data such as epidemiological studies might be of importance).

- For evaluating reliability of studies in the REACH and biocides area (also the RAC methodology for setting OELs refers to the respective guidance) Klimisch scores (Klimisch et al. 1997) are used; no specific criteria are mentioned by others. Similarly, the use of read-across and QSAR methods is described in detail in the documents relevant for REACH and biocides only.

- It can be assumed that committees consider it less important to provide detailed guidance on steps like data search and quality evaluation, as there might be a mutual understanding among committee members on how to perform these activities; however, for the sake of transparency and long-term consistency of evaluations, procedures laid down in writing would be beneficial.

- In all systems human data are regarded as relevant information source, but from practical experience it is evident that more weight is given to that kind of information (including short-term experiences of workers) by workplace committees, whereas in the REACH, PPP and biocides area much weight is given to the experimental data forming part of the submissions. In addition, OELs are often derived for data-rich substances, for which availability of reliable human data it is more likely.

- Justifications for substance-specific OELs are documented in detail; the existing examples provide respective insight, although requirements are not detailed in the method documentation; it is noted that quality of substance-specific documentation provided by the MAK commission, AGS and SCOEL improved substantially over time. For REACH, PPP and biocides the regulatory process defines how the value derivation is reported, but details on the derivation of substance-specific REACH DNELs are often not in the public domain.

2.3 Methodology for deriving limit values for systemic effects

In the following the individual steps leading to a numerical OEL setting are discussed, among them

- determination of a point of departure (POD)
- consideration of differences in exposure conditions between humans and experimental animals
- use of uncertainty/assessment factors.

The general methodologies for deriving OELs for systemic effects are described here. Specific considerations for OELs for local effects in the respiratory tract are discussed in section 2.4.

2.3.1 Observations

The following table summarises key observations made in the detailed analysis as documented in Annex 1.

Table 2-3 Observations on individual steps of deriving OEL (-analogue) values

Subject	Observation
Applicable POD type	All systems use NOAELs (NOAECs) or benchmark doses as POD, some in addition allow to use LOAELs as potential POD; few method documents are specific with regard to use of the BMD or BMDL and the benchmark response level to be used for setting a BMD/L
Selection of POD	Several documents state that the POD leading to the lowest OEL (-analogue) value should be used, whereas others mention that the "most sensitive species" or the "most sensitive endpoint" should be used
Route-to-route extrapolation	In principle, possible, but all systems formulate reservations or specific criteria for route-to-route extrapolation
Modification of POD and anthropometric data	In all systems provisions are made for modifications of the POD for differences in absorption, exposure patterns or physical activity. But the level of details explaining how to perform it varies substantially. For AGS practical experience shows that provisions from the guidance on exposure-risk relationships for carcinogens (AGS 2013) are also used for deriving OELs; few information is also provided on default anthropometric data and respective data for the experimental animal (body weights, inhalation rates, etc.)
Use of assessment factors (AF)	In principle, AF are used in all systems analysed; however, several relevant differences are obvious (see Table **2-5**); major differences are - MAK commission and SCOEL do not prescribe use AF for time extrapolation (although in practice MAK commission uses AGS factors, modified on a case-by-case basis*); factors used in all other systems are largely identical - SCOEL does not provide default values for inter- and intraspecies extrapolation; MAK commission applies a factor of 2 for inter-/intraspecies extrapolation - the intraspecies AF varies between 1 and 10 in the various systems - in all systems but for pesticides and biocides allometric scaling with exponent 0.75 is used as a first step for interspecies extrapolation; the factor for (remaining) interspecies variability varies largely - extrapolation from a LOAEL is not recommended by most, with reference to the benchmark dose method
Deviation from default values	Deviation from defaults is possible in all systems, e.g., by use of PBPK models or chemical-specific adjustment factors (CSAF)
Other AF	Additional factors for the severity of effects or the (poor) quality of the overall database might be used in some systems, e.g., for pesticides and biocides, but are typically not applied by workplace committees

* Personal communication, MAK Commission, December 2020

2.3.2 Conclusions

Differences in the approaches used were identified, which might give rise to quantitative differences in derived values. The most relevant are:

- The guidance documents are not always clear about whether the key study and POD should be used, which yields the lowest OEL or whether the lowest POD should be used (this might often, but not always be the same, especially if both human and animal data are available.

- NOAEL and BMD/L are the preferred types of PODs, but practical examples indicate that LOAEL-NAEL extrapolation is still often used (for example, see AGS 2015).

- Modification of the POD to consider differences in exposure conditions, physical activity and/or absorption between the experimental animal and workers is another source of differences (see **Table 2-4** for details).

- A major source of quantitative differences is the AFs for time extrapolation (variation from 1 to 6) and intraspecies extrapolation (variation from 1 to 10); **Table 2-5** summarises the defaults proposed.

- An obvious difference exists with regard to interspecies extrapolation: allometric scaling (plus a factor to account for the remaining uncertainty) is used in all systems, but the PPP and biocides area; here, a default factor of 10 is applied, as it is still in use, e.g. in the evaluations of WHO in the food area (WHO 2009)

There is a clear difference in the use of (default) assessment factors between OEL-setting workplace committees (SCOEL, MAK commission) and other regulatory areas, which becomes most obvious for time and intraspecies extrapolations: OEL committees tend to avoid default factors (or set it to 1) and prefer to conclude on quantification of extrapolation steps case by case.

Table 2-4 Provisions for modifying the POD

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
Correction for differences in route-specific absorption	Yes, If no substance specific data is available, the following defaults apply: Oral-to-inhalation: 50% oral absorption, 100% inhalation absorption. Inhalation-to-oral: 100% inhalation absorption, 100% oral absorption.	Not mentioned	Yes, If no substance specific data is available, the "allometric scaling" procedure for oral-to-inhalation extrapolation, from R.8 is followed, but it is explicitly noted that no default for oral absorption fraction is used. Instead, the absorbed fraction is discussed in a case-by-case basis. The	Yes, If no substance specific data is available, complete absorption via both routes is assumed	Yes, If no substance specific data is available, the "allometric scaling" procedure for oral-to-inhalation extrapolation, from R.8 is followed, but for oral absorption a default of 100% (instead of 50%) is used. For metals and metallic compounds a default of 50% absorption	Yes, Does not propose own defaults, due to considerable scientific uncertainty. Refers to R.8 (50% oral absorption and 100% inhalation absorption), but recommends to generate substance specific data.	Yes, No default for oral absorption, but 100% is implied Default for respiratory absorption: 100%	Yes, If no substance specific data is available or available data does not indicate an absorption significantly below 100%, a default value of 100% for oral absorption should be used. For inhalation no defaults are provided.

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
			default for inhalation absorption is 100%.		applies. The default for inhalation absorption is 100%.			
Correction for different exposure conditions (e.g. exposure duration)	Yes, Defaults for workers: 8h daily exposure, 5 days/week 48 weeks/year, 40 years/life	Yes, Refers to SCOEL methodology 2017: 40 hours/week, 48 weeks/year, 40 years/life	Yes, 40 hours/week, 48 weeks/year, 40 years/life	Not mentioned	Yes	Yes, refers to R.8: 8 h daily exposure, 5 days/week, 48 weeks/year, Working lifetime not explicitly mentioned.	Yes 8 h daily exposure of workers is given as a default, but no other defaults for workplace exposure times.	Yes, including different time patterns of exposure 8 h daily exposure of workers is given as a default, but no other defaults for workplace exposure times. Refers to R.8 for further guidance.
Correction for different	Yes,	Yes,	Yes,	Not mentioned in resp.	Yes, for gases and vapours	Yes,	No	Yes,

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
physical activity	10 m³ human respiratory volume under 8 h light activity, 6.7 m³ human respiratory volume under 8 h of rest	Refers to SCOEL methodology 2017: the assumed respiratory volume of a worker is 10m³/8 h	the assumed respiratory volume of a worker is 10m³/8 h. This is assumed to be twice as high as the respiratory volume of an experimental animal over a 6 h duration.	guidance, but in practice AGS (2013) is followed; assumes 10m³/8 h for slight physical activity	with blood:air-partition coefficient > 5 and only for systemic and lung effects. The assumed respiratory volume of a worker is 10m³/8 h. The resulting dose/kg is assumed to be twice as high as the dose/kg for an experimental animal over a 6 h duration.	refers to R.8: 10 m³ human respiratory volume under 8 h light activity, 6.7 m³ human respiratory volume under 8 h of rest		Refers to R.8 for further guidance. However, for assessing systemic exposure, recommended values from the US EPA Exposure Factors Handbook (US EPA 1997) are reported, which differ slightly from R.8

Table 2-5 Default assessment factors

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
Default AF for time extrapolation	sa* – c: 6 sa – sc: 3 sc – c: 2	Yes; refers to R.8.4.3: sa* – c: 6 sa – sc: 3 sc – c: 2	No	sa* – c: 6 sa – sc: 2 sc – c: 2	Mentioned only for irritation (see below), although in practice factors are applied also for systemic effects sa* – c: 6 sc – c: 2	sa* – c: 6 sa – sc: 3 sc – c: 2	sa* – c: - sa – sc: - sc – c: 2	sa* – c: 6 sa – sc: 3 sc – c: 2
Allometric scaling for interspecies extrapolation	Yes, exponent 0.75	Yes, exponent 0.75	Yes, exponent 0.75	Yes, exponent 0.75	Yes, exponent 0.75	Yes, exponent 0.75	No	No, but can be used to replace default AF
Default AF for interspecies extrapolation	2.5	2.5	No default provided	Inter + Intra = 5	Typically, OEL set at 50% of extrapolated NOAEL (i.e., combined factor 2 for intra- and	1	10	10

	REACH Regulation	RAC OEL metho-dology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	Plant Protection Products Directive	EU Biocidal Products Regulation
					interspecies extrapolation)			
Default AF for intraspecies extrapolation	5	5	>=1	Inter + Intra = 5	Typically, OEL set at 50% of extrapolated NOAEL (i.e., combined factor 2 for intra- and interspecies extrapolation)	3	10	10
AF for severity of effects	1, should be increased on a case-by-case basis	1, should be increased on a case-by-case basis	Not explicitly addressed	Not considered necessary	Not explicitly addressed, considered on a case by case basis	1, requires larger AF if severe effects at LOAEL	≤ 10 on a case-by-case basis, (e.g. for teratogenic or irreversible neuropathic effects)	2 -10, taking into account dose-response data
AF for LOAEL-NAEL extrapolation	Yes; 3 – 10	Yes; 3 – 10	Use NOAEL or BMD	Use NOAEL or BMDL, although in practice on a case-by-case basis also LOAELs are	Not mentioned; however, factors given for irritating effects are applied in practice also	3	Use NOAEL or BMD, no default	Use NOAEL or BMD, no default

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
				used as POD (with factor 3)	for systemic effects			
Quality of database	1, should be increased on a case-by-case basis	1, should be increased on a case-by-case basis	Yes, no default provided	Not mentioned	Not mentioned; however, in practice quality of data base is considered by expert judgment	1, should be increased on a case-by-case basis	Yes, no default provided	Yes, no default provided

*: sa: subacute, sc: subchronic, c: chronic

2.4 Methodology for deriving limit values for local effects on the respiratory tract

In the same way as for systemic effects, the individual steps leading to an OEL for local effects in the respiratory tract are compared, such as

- determination of a point of departure (POD)
- calculation of a human equivalent concentration
- use of uncertainty or assessment factors.

The documentations vary in the depth in which they discuss local effects. For example, no specific consideration is given to local effects in the ECHA guidance on deriving OELs or in the guidance documents on biocides and PPP. It is assumed that general principles as laid down for systemic effects apply also for assessment of local effects and are not repeated.

The most relevant local effects are sensory irritation and respiratory toxicity in the upper and lower respiratory tract. Note that for effects caused by particles (aerosols) in the lower respiratory tract a separate report is under preparation (Part 5: Effects of aerosols in the lower respiratory tract).

2.4.1 Observations

The following table summarises key observations made in the detailed analysis as documented in Annex 1.

Table 2-6 Observations on individual steps of deriving OEL (and analogue) values for local effects in the respiratory tract

Subject	Observation
Applicable POD type	All systems use NOAELs (NOAECs) or benchmark doses as POD, some in addition allow to use LOAELs as potential POD
Modification of POD and anthropometric data	There are only a few documents, which spell out specifically how local effects should be evaluated. In these, similar provisions as for systemic effects are proposed for differences in the exposure scheme and for physical activity. But the application of the modifications is restricted to substances, whose effects are not purely concentration-dependent
Use of assessment factors (AF)	Again, large differences exist in default AF. ECETOC recommends a default factor of 1 for time extrapolation for local effects (assumption of concentration-dependency under all circumstances), in contrast up to a factor 6 for subacute to chronic extrapolation is used in several other systems; for intraspecies extrapolation there are again large differences with defaults ranging from 1 to 10; for an overview see **Table 2-8**
How is sensory irritation considered?	Sensory irritation is discussed as an important endpoint for consideration by workplace committees (although it is not mentioned in the new ECHA guidance on how RAC will derive OEL values) and specific default AFs are used in Germany for this endpoint (see below). According to REACH guidance, experimental data on sensory irritation (Alarie test) should be used for short-term DNELs only. Sensory irritation is not mentioned in the guidance for pesticides and biocides
How is deposition and clearance of aerosols in the respiratory tract considered?	Evaluation of deposition and clearance of aerosols in the lower respiratory tract can be another reason for divergence. In Germany, this is explicitly considered for setting OELs by AGS and models such as MPPD are used where possible. MPPD is also recommended by ECETOC. Other organisations acknowledge potential species differences in deposition and clearance but consider these to be covered by the interspecies AF and do not provide specific recommendations.

2.4.2 Conclusions

The emphasis given to local effects in the airways clearly differs between the regulatory areas. Much less weight is given to local effects in the area of pesticides and biocides, as these are mainly deriving AOELs or AELs as systemic absorbed doses.

- Where addressed, similar POD modifications are recommended as for systemic effects (without adjustments for differences in absorption).

- For local effects inhalation studies are used and the POD is typically expressed as air concentration. Hence, no allometric scaling needs to be applied.

- Large differences exist in proposed default AF values, especially for time and intraspecies extrapolation (see Table 2-8).

- Sensory irritation is discussed and acknowledged as relevant endpoint by the workplace committees, but gains less attention in other regulatory areas. In Germany, specific extrapolation factors are in use for sensory irritation as the basis for deriving an OEL (Brüning et al. 2014) by AGS (although not specifically mentioned in the guidance) and MAK commission; these default AF are shown in **Table 2-7**; they will be further discussed in the context of the individual extrapolation steps later in the project.

- Differences in deposition and clearance of (solid and liquid) aerosols in the lower respiratory tract between experimental animals and humans are explicitly considered currently (by using deposition models such as MPPD and procedures to calculate a human equivalent concentration, HEC) only for German OELs. For aerosols this might lead to quantitative differences in the assessment.

Table 2-7 Default assessment factors for sensory irritation according to Brüning et al. (2014)

	Factor	**Remarks**
Time extrapolation	6 2 1	from subacute experimental study to chronic human conditions from subchronic experimental study to chronic human conditions Comparison between (high quality) human controlled experimental study and chronic exposure under workplace conditions
Interspecies extrapolation	3 approx. 2 - 3	from chronic inhalation NOEC (animal study, typically with histopathological confirmation of effects in upper airways) to human NOAEC for sensory irritation if olfactory epithelium is target (case-by-case consideration: “It should be considered to reduce the default iEF to 2”)
Intraspecies extrapolation	1	“if OEL is derived from human sensory NOAEC since it is based on a controlled human exposure study” no default proposed for derivation based on animal study*
LOAEL - NAEC	2 or 3	depending on severity of effects and steepness of dose-response relationship (DFG 2018)

*No recommendation for consideration of human variability, if OEL derivation is based on animal study, is available; in practice, 1 is also default for derivation based on animal study.

Table 2-8 Default assessment factors for local effects in the respiratory tract

	REACH Regulation	**RAC OEL methodology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
Default AF for time extrapolation	sa* – c: 6 sa – sc: 3 sc – c: 2 (lower, if effects are concentration-dependent)	-	No explicit explanation for local effects given	sa* – c: 6 sc – c: 2	sa* – c: 6 sc – c: 2 (for sensory irritation, in practice also used for other effects)	1 for all time extrapolations	-	Reference to REACH assessment factors is made
Allometric scaling for interspecies extrapolation?	No – assessment based on concentration	-	No – assessment based on concentration	No – assessment based on concentration	No – assessment based on concentration	No – assessment based on concentration	-	No
Default AF for interspecies extrapolation	2.5	-	Specific correlation for sensor. irritation (Alarie) in annex	Inter + Intra = 5	2 – 3 (for sensory irritation only)	1	-	2.5 (based on air concen-tration)
Default AF for intraspecies extrapolation	Worker: 5	-	2 for sensor. irritation, no default for other local effects	Inter + Intra = 5 (see above)	1	Worker: 3	-	10 (for professionals and non-professionals)
AF for severity of effects	1, should be increased on	-	No explicit explanation for	Not mentioned	Not explicitly addressed, considered	No specific provisions for local effects	-	Reference to REACH

	REACH Regulation	RAC OEL methodology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	Plant Protection Products Directive	EU Biocidal Products Regulation
	a case-by-case basis		local effects given		on a case-by-case basis			assessment factors
AF for LOAEL-NAEL extrapolation	Yes, 3 - 10	-	2 – 3 for sensor. irritation, no default for other local effects	LOAEL not foreseen as POD	2 - 3 (for sensory irritation only, in practice also used for other effects)	No specific provisions for local effects	-	Reference to REACH assessment factors
Quality of database	1, should be increased on a case-by-case basis	-	No explicit explanation for local effects given	Not mentioned	Not mentioned, in practice considered by expert judgment	No specific provisions for local effects	-	Reference to REACH assessment factors

*: sa: subacute, sc: subchronic, c: chronic

2.5 Carcinogens (with thresholds)

This comparison of methods focusses on substances, for which OELs or OEL-analogue values are derived assuming existence of thresholds. For (genotoxic) carcinogens the threshold concept typically is not applied. Instead, exposure-risk relationships are derived, which inform about the expected excess risk at a specified exposure level (AGS 2013; ECHA 2012). In the REACH context a derived minimal effect level (DMEL) needs to be calculated for non-threshold carcinogens, using the exposure-risk relationship or other methods described in the ECHA Guidance document R.8 (ECHA 2012).

For some carcinogens, however, it is assumed that they act via threshold-like modes-of-action. For these substances OELs might be derived in a similar way as discussed in the chapters above. Based on the conclusions from the Joint Task Force of SCOEL and RAC the RAC methodology for deriving OELs state that the no-threshold assumption is the default approach, but states that "when subsequent analysis of the data allows refinement in the sense that overall the data actually points to a threshold, then a threshold approach can be followed. Without (sufficient) data to conclude this, the default stays a non-threshold MoA" (ECHA 2019). But no guidance is given on the level of evidence required to deviate from the default approach of deriving an exposure-risk relationship. Recently, RAC derived OELs for benzene, nickel and nickel compounds and acrylonitrile (RAC 2018a, b, c). In all three cases a mode-of-action based threshold was assumed. These practical examples provide insight into how RAC is dealing with such cases, when deriving OELs for inclusion in the occupational safety and health legislation. Different types of endpoints were chosen in these cases for finding the POD: genotoxicity in the case of benzene, respiratory and developmental toxicity for nickel compounds and carcinogenic effects in the case of acrylonitrile. Assessment factors (including an extra factor for the severity of carcinogenic effects) were applied to the POD to derive the OELs.

SCOEL did not derive OELs for

- genotoxic carcinogens (Group A) or for
- substances for which direct genotoxicity cannot be excluded due to missing or incomplete data (Group B)

In such cases, "cancer risk estimates at various exposure levels may be calculated" (SCOEL 2017). For specific cases, classified either as

- weakly or indirectly DNA-reactive (Group C), or
- non-genotoxic carcinogens (Group D),

"a "true" or "practical" threshold may be derived which protects from carcinogenicity or an extra cancer risk, for example by preventing inflammation or accelerated cell division due to irritation" (SCOEL 2017). These true or practical thresholds of SCOEL can be considered to be equivalent to RAC's "mode-of-action based thresholds". But practical examples (see RAC's opinions on benzene, nickel and nickel compounds and acrylonitrile (RAC 2018a, b, c)) show that RAC is proposing to use additional assessment factors for the severity of effect, leading to quantitative differences compared to SCOEL recommendations.

In Germany, in a similar way, a mode-of-action analysis is applied for carcinogens and either an exposure-risk relationship or an OEL is derived by AGS. In the respective guidance document it is explained that detailed information is required with regard to the mode(s) of action active in a specific case. Primary and secondary genotoxicity is discussed as well as other modes of action. It is emphasised that several mechanisms might be active in parallel. In consequence of the mode-of-action discussion either

- a linear exposure-risk relationship
- a sublinear exposure risk relationship or
- a threshold-based OEL.

is derived. Detailed guidance is given how to quantitatively derive linear or sub-linear exposure-risk relationships. For threshold-based OELs severity of effect is considered by applying an additional assessment factor.

In the methodology followed by the MAK commission an OEL is possible for carcinogens in category 3 and 4, provided that genotoxicity is not a prime concern. For category 5 carcinogens a MAK value is defined as bearing a very low carcinogenic risk.

3 Discussion

3.1 Definition and scope

Definitions

In principle, all values considered have similar definitions. They aim at identifying doses or concentrations, at or below which no detrimental effects are expected (Table A-1). In conclusion, all OEL frameworks discussed here are leading to health-based values. This is not always the case with legally established OELs (e.g. BOELV in the European Union), which – in addition to health considerations – might be policy-driven or influenced by technical requirements (Deveau et al. 2015; Waters et al. 2015).

An obvious difference exists in the units, in which values are presented: in contrast to all other systems AOELs and AELs for plant protection products and biocides, respectively, focus on the internal systemic dose, whereas all others provide OELs and workers DNELs as air concentrations. This is symptomatic for the larger weight, which is given on systemic effects in the PPP and biocides area, and the requirement there to consider the joint contributions of all pathways to systemic exposure.

Inclusion of sensitive individuals or groups and reproductive toxicity endpoints

All methodologies claim to include sensitive individuals or subpopulations (without being precise with regard to e.g. quantiles of the total working population or consideration of especially susceptible individuals such as asthmatics or nickel sensitive persons).

It is noteworthy that obviously the methodology of SCOEL changed in several fundamental points from 2013 to 2017 (only the 2017 guidance document was considered in the tables in Annex 1). In the 2013 guidance document it is stated that “OELs are established for healthy workers” (SCOEL 2013), which has to be interpreted as OELs not guaranteeing protection for workers of poor health. In contrast, in the 2017 methodology paper “existing disabilities or underlying disease” is explicitly acknowledged as a cause of variability in the working population (SCOEL 2017). Similarly, although their OELs are meant to limit exposure so that it “will not lead to adverse effects on the health of exposed persons and/or their progeny”, in the 2013 guidance document SCOEL stated with regard to the inclusion of women of childbearing age: “OELs established to protect adults cannot a priori guarantee the absence of pre- or post-natal adverse effects. Thus pregnant or lactating women may represent a special risk group in the workplace” (SCOEL 2013). In the 2017 guidance document it is stated that OELs are regarded “adequate to protect the health of the workers (and the health of their offspring, as regards developmental effects that may be caused by chemical agents” (SCOEL 2017). Consequently, previously derived OELs may not be in line with this (new) definition.

With regard to the inclusion of developmental toxicity there seems to be a principal difference between OELs as set by SCOEL, RAC or German committees and the

values derived in other legislative areas (REACH, biocides, plant protection products): The latter require the quantitative consideration of this endpoint for setting values.

In Germany, a quantitative comparison is done which might lead to a notation that the OEL is capable/not capable to protect women of child-bearing age: The NOAEL for developmental toxicity divided by 10 is compared with the OEL based on other effects. Taking into account the factor of 5 used by AGS (2010) this might be interpreted as an addition factor of 2 for developmental effects. Reproductive toxicity is also considered by MAK Commission. For substances of pregnancy risk group B a concentration may be given which would correspond to a classification in pregnancy risk group C (Damage to the embryo or foetus is unlikely when the MAK value or the BAT value is observed.)

The situation is less clear for RAC and SCOEL. Whereas RAC (ECHA 2019) states: “Because of the relative sensitivity of the rapidly developing individual (from conception to puberty) to specific toxic effects, OELs established to protect adults cannot a priori guarantee the absence of pre- or post-natal adverse effects.”) and SCOEL’s previous methodology (SCOEL 2013) seem to not include protection of women of child-bearing age, there is a change of position in the new SCOEL guidance (see discussion in previous paragraph).

In a discussion of SCOEL and RAC on differences between the SCOEL OEL and the RAC DNEL for workers for N-methyl-2-pyrrolidone both parties agreed that respiratory irritation, (as evidenced by chemosensory effects) as well as developmental effects can be used to derive such values (RAC/SCOEL 2016). However, they failed to come to a common position regarding the most relevant endpoint and the determination of a POD. From this example one can conclude that developmental toxicity is considered a relevant endpoint for deriving OELs by both parties. But it is noted that the DNEL for N-methyl-2-pyrrolidone was derived by RAC in the context of a restriction proposal and it remains to be seen whether developmental toxicity data are considered the same way when deriving OELs according to the new guidance document (ECHA 2019).

With regard to fertility, from existing evaluations it can be concluded that effects on male fertility are considered in all systems. Only AGS (2010) provides specific extrapolation factors to be used for male fertility, whereas no empirical investigations (see Schuhmacher-Wolz et al. 2006) and no specific recommendations for effects on female fertility exist.

Inclusion of local irritation and respiratory sensitisation

A further difference between the legislative areas of biocides and plant protection products with workplace-specific applications is in the importance given to local effects in the respiratory tract, as seen from the inclusion of sensory irritation data. Such data and effects are often used to derive OELs, whereas it is obvious from the description that this is not a major consideration for plant protection products and biocides. Similarly, an analysis performed by Schenk and Johanson (2019) indicates that respiratory sensitisation of substances is often not considered by registrants for deriving worker DNELs, but is taken into account (to some extent) in some of the OELs. Examples studies were trimellitic anhydride, phthalic anhydride and maleic anhydride.

DNELs were orders of magnitude higher than OELs proposed by the US American Conference of Governmental Industrial Hygienists (ACGIH), most likely because respiratory sensitisation was not considered when deriving the DNELs. But it should be noted that no generally accepted methodology exists to derive OELs based on allergic effects in the respiratory tract. The MAK commission states that no validated experimental model exists to investigate respiratory sensitisation in the animal (DFG 2018).

The discussion on local effects in the respiratory tract will be further expanded below.

Short-term values

In addition to time-weighted average concentrations to assess long-term exposures, setting OEL values typically involves consideration of the relevance of short-term exposures. If a substance shows high toxicity after short-term exposure, short-term OELs are set by a fixed value or by a x-fold exceedance value. For example, the German TRGS 900 defines exceedance factors (max. 8), which are valid for exposure periods of 15 minutes, which should not occur more often than 4-times per shift and with an interval of 1 hour (AGS 2018). Similarly, SCOEL methodology foresees STEL values for a 15 minute-exposure period and a maximum frequency of 4-times per shift and intervals of 1 hour.

Acute inhalation DNELs for workers can be derived under REACH, if the toxicity and exposure profile requires this. But, as explained in an own appendix in the guidance document R.8 on the topic, this might become difficult when having to rely on experimental data only. Acute AOELs and AELs are typically derived for oral exposure only.

The following table highlights some differences between workplace-specific systems and the procedures in other regulatory areas.

Table 3-1 Differences between “OEL systems” and other regulatory areas

consideration of	**“OEL systems”**	**REACH DNELs**	**AOELs (PPP) and AELs (biocides)**
developmental toxicity	no/? (see text)	yes	yes
male/female fertility	yes/?	yes	yes
endocrine disruptors	No*	(yes)	yes (for hazard identification)
sensory irritation	yes	for short-term values	no
respiratory sensitisation	notation (on a case-by-case also for quantification)	qualitative assessment only	not specifically mentioned (PPP) or qualitative assessment
short-term values	STEL (15 min)	acute DNEL	AAOEL / AAEL

* although not mentioned, in practice endocrine disrupting activity may often be considered

Assessment of endocrine disrupting effects is pivotal for active substances in plant protection products and in biocidal products, as endocrine disruptors are not allowed to be used in these products. Therefore, it is not surprising that weight is given to this endpoint in the hazard evaluation of endocrine disruptive properties. But note that identification of a substance as endocrine disruptor will in most cases lead to its rejection for use in biocidal or PPP products. So, the data are used for hazard identification, but not for deriving A(O)ELs.

Also under REACH this endpoint is increasingly given emphasis, as endocrine disrupting chemicals might be assessed as substances with “equivalent level of concern” under Art. 57 f), and consequently included in REACH Annex XIV. Endocrine disruptors are not specifically mentioned in the guidance documents on deriving OELs.

It should be noted that there are policy decisions, which also impact the methodological approaches, such as

- Non/-inclusion of unpublished studies
- Non/-inclusion of risk management aspects, such as technical feasibility or analytical detection limits
- overall protection goals.

Methodologies are only comparable, if they agree in these pre-set conditions. Transparency on these conditions is therefore mandatory.

Recommendations

- Methodological descriptions should include a description on how fertility effects and developmental toxicity is considered in the derivation of values and whether women of child-bearing age and their progeny are expected to be protected by the derived values

- The obviously existing differences in that matter ask for harmonisation

- All methodologies should explicitly state how assessors should deal (qualitatively or quantitatively) with local respiratory effects (including sensory irritation) and respiratory sensitisation.

3.2 Data search and evaluation

Databases

In the investigated documents the level of detail on information sources and requirements for data retrieval and evaluation varies substantially. Where detailed descriptions are available most organisations agree on a careful evaluation of data quality and apply weight-of-evidence approaches, although the latter ones are not explained extensively (see Table A-2).

Guidelines established for systematic reviews may help to define requirements for data searches and its documentation (see e.g. Liberati et al. 2009). However, systematic review requirements are difficult to apply directly to toxicological risk assessments aiming at a complete overview on the toxicity of a substance, which is required for deriving an OEL. According to Hoffmann et al. systematic review methods have been developed for answering specific, narrow primary research questions (Hoffmann et al. 2017). An example for such a specific question, which has been answered by a systemic review is whether an association exists between cadmium exposure and prostate cancer (Ju-Kun et al. 2016). For broader questions the specific requirements of systematic reviews for data evaluation and detailed documentation of decisions on all records found might become very laborious and resource-intensive. Hoffmann et al. identified further challenges such as availability of information specialists providing systematic review experience and being sufficiently familiar with peculiarities of toxicological data; establishment of processes to publish, discuss and agree on search protocols, and handling vast amounts of records (Hoffmann et al. 2017). Nevertheless, criteria and procedures developed for systematic reviews can, if adapted to the specific needs, help to establish transparent guidance for data search and documentation requirements also for reviews aiming at deriving OELs.

Animal and human data

In principle, all methodologies include use of human and animal data, although some (e.g. SCOEL) explicitly mention that priority is given to high quality human data. So, no substantial difference should be expected in data selection. However, an empirical investigation by Schenk and Johanson showed that bodies engaged with deriving OELs more often used human data than, for example, REACH registrants when deriving DNELs: out of 75 IOELVs analysed by Schenk and Johanson (2010), 31 values were based on human data. This is at least partly explained by the fact that OELs are derived for fewer, more data-rich substances (for which availability of relevant human data is more likely) compared to substances requiring registration under REACH. But also a general preference for human data over experimental animal data on the side of OEL-deriving committees might be contributing to this observation. For example, SCOEL explicitly explains that “human studies with populations encompassing workers are more relevant than animal studies” (SCOEL 2017). Based

on the findings of the Joint RAC/SCOEL task force (ECHA/RAC-SCOEL 2017a) in RAC's documentation on OELs preference is given to "using good quality human data when available" (ECHA 2019). But the challenge for a harmonised approach will be to define "good quality" and how to evaluate the respective dis-/advantages of human versus experimental data.

Schenk identified differences in the database used as a possible reason for differences in OELs derived by different bodies (Schenk 2010). It was also noticed that e.g. SCOEL in former years based its evaluations on existing reviews and assessments and consulted original studies only to a limited extent (Schenk and Johanson 2018). It can be concluded that the implicit or explicit criteria for selecting key information for the assessment might be a reason for differences. Lavelle et al. proposed a detailed framework for evaluating human and animal data to facilitate the decision on the most relevant information to base the assessment on (Lavelle et al. 2012). From the analysis it became clear that detailed guidance is necessary for searching, selecting and evaluating data. A transparent quality evaluation of key studies could indeed help to avoid discrepancies between assessments. At the same time, documentation requirements should be balanced.

Application of QSAR and read-across approaches seem to be possible in most, if not all, systems. However, the level of explanation varies from simple mentioning the terms to extended guidance documents. Typically, the workplace-oriented systems are not very explicit with regard to the criteria for applying these methods.

Documentation

Documentation requirements for REACH chemicals, plant protection products and biocidal products are obvious due to the role of the assessments in the regulatory process (i.e. submissions have to be made to the regulatory authorities and the formats for reporting are defined). In contrast, no description of documentation requirements exists for OEL justifications from expert bodies. But all substance-specific evaluations are published and the form and detail can be deduced from these publications.
Harmonising documentation of OEL derivation was proposed as a means to harmonise values and to increase transparency and comparability (Deveau et al. 2015). An increase in transparency can be achieved by comparing the derived values with the previous ones and those of other institutions existing in parallel in order to explain differences and changes.

Rules for updates

It is noteworthy that, although it is common sense that scientific evaluations need to be adapted in case new information becomes available, in general there is no information in the method descriptions on how an update is triggered (e.g. on a regular basis or upon availability of new data). Update requirements exist in principle for all REACH dossiers, and, hence, also for DNELs, when new relevant information becomes available. The German committees publish lists of substances subject to (re-)evaluation, however, without transparent criteria for inclusion in those lists.

Recommendations

Methodological descriptions should include

- a description of the type of data to be searched, the range of sources (bibliographic and other databases, grey literature) and requirements for documenting search criteria and results
- a procedure to evaluate the quality and suitability (reliability and relevance) of data, while keeping the workload manageable
- a description of how weight-of-evidence should be used (major considerations in a weight-of-evidence approach)
- considerations and criteria on the use and weight of human and animal data
- information on criteria for and frequencies of assessment updates.

3.3 Methodological steps for deriving OELs for systemic effects

Availability of guidance

The provided guidance shows large differences:

There is detailed guidance available on deriving DNELs and AELs for REACH registrations and biocidal products. This is not surprising, as these documents were specifically prepared for use by third parties (registrants, submitters). Incomplete guidance is available for deriving AOELs in the context of plant protection products. The process of developing such a guidance was discontinued, with the effect that various documents need to be consulted and major gaps exist in the description of a step-by step procedure to derive AOELs.

With the revised SCOEL guidance a consistent document is available covering many aspects of the derivation process. The draft guidance on how RAC will derive OELs refers to ECHA's REACH-related guidance as well as to the SCOEL document and the reports of the RAC/SCOEL task force. It would be advantageous, if this could be reorganised as a stand-alone document specifically addressing all aspects.

The two German approaches provide less guidance compared to the ones described above. Development of comprehensive documents to explain how MAK values as well as AGW values of TRGS 900 are to be derived would be helpful.

It should be noted that procedures to derive OELs or OEL-analogue values undergo changes, which are not always reflected in the method descriptions. For example, the German MAK commission increasingly considers assessment factors, as used by AGS, in their evaluations, although this changing practice is not reflected in the (short) methods documentation. On the other hand, (older) OEL are not always in line with

recent changes in methodology. OELs derived in the past by SCOEL would need to be updated to make them concordant with the new guidance document (SCOEL 2017).

Determining the point of departure

The analysis of the individual methodological steps shows areas of agreement as well as differences in the approaches (see section 2.3 and Annex 1). For example, LOAEL – NAEL extrapolation is an option available for REACH chemicals (and, as RAC refers to the respective guidance also for RAC OELs), plant protection products and biocides, whereas SCOEL and the German AGS claim to use NOAEL and BMD as point of departure only (although in practice also LOAELs are used as POD on a case-by-case basis).

The REACH Guidance and ECETOC specify the BMDL05 as point of departure (POD) for quantal data equivalent to a NOAEL. In contrast, SCOEL usually uses the BMD05. All others are not specific as to which benchmark rate (BMR) and whether the central estimate (BMD) or the lower bound of the confidence interval (BMDL) should be used. None of the guidance documents are recommending a BMR for continuous data. In general, very little guidance is given on the application of the benchmark dose. Wheeler et al. (2015) clearly advocate for using dose-response modelling for deriving OELs, but the guidance documents evaluated clearly indicate the current practice that many systems consider this a valid alternative at most and use of the NOAEL is probably still the most often used approach.

A further source of uncertainty is the criteria for determining the key study (or studies) from which the POD is taken: some methods indicate that the study/POD leading to the lowest final value should be used, whereas others recommend using data from the most sensitive species or endpoint (which can be interpreted as the lowest NOAEL). Although this is expected to lead to the same choice in many cases, there might be situations where different conclusions result, if e.g. human data and experimental data are compared based on NOAELs or assessed in parallel down to the resulting values. The Joint ECHA/RAC/SCOEL Task Force (2017a) identified the selection of the POD as a step potentially leading to numerical differences between SCOEL and RAC-derived values and asked for more guidance on this topic.

In an analysis of existing OELs and their justifications as derived by various national OEL committees, Schenk (2010) concluded that even within the same regulatory field large quantitative differences may occur. Determination of the key studies and the POD was an important reason for differences (caused by differing views on study quality and critical effects). Schenk et al. (2015) compared DNELs for workers derived by registrants with Swedish OELs and DNELs derived by the author according to their interpretation of the guidance document R.8. Registrants DNELs on average were similar to Swedish OELs, but higher than DNELs derived by the authors. Variability was huge in all sets compared, and the authors identified choice of key studies, POD and assessment factors all contributed to the differences. They concluded that despite the extensive REACH guidance there are still many choices to be made without detailed instructions, especially for deviations from defaults. One example was selection of a suitable time extrapolation factor for developmental toxicity effects. With

regard to this example it should be mentioned that the German AGW system provides specific assessment factors for male fertility and concludes that, based on adequate studies for this endpoint, no time extrapolation is required (AGS 2010). This can be seen as a further refinement step, but might also lead to differences between approaches, if not taken up by others.

Use of assessment factors to account for uncertainty and variability

Principal differences are obvious in the surveyed systems with regard to the use of assessment factors (see **Table 2-5**). All organisations agree that chemical-specific information should be used, whenever available, to reduce uncertainties and to set adequate extrapolation factors to account for remaining uncertainty and variability. But large differences exist with regard to quantification of extrapolation factors in case substance-specific information is absent. In the PPP and biocides area the "traditional" 10 x 10 safety factor approach is still used, with additional factors e.g. for time extrapolation or severity of effects. All others use the allometric scaling approach (metabolic rate scaling with exponent 0.75) and an intraspecies variability factor of <10 for workers. But several numerical differences exist also in the factors used by these organisations

Large differences were observed for time extrapolation, with factors varying from 6 for subacute to chronic to not using AFs for time extrapolation at all. Further, intraspecies variability seems to be a relevant reason for numerical differences. In the surveyed systems the applied factors reach from >= 1 to 10 or are set on a case-by-case basis. Traditionally, OELs were set without explicit consideration of genetic susceptibility or other sources of inter-individual variability, as analysed by Schulte and Howard (2011). Johannson et al. reported that susceptibility of asthmatics was often not taken into account by registrants when deriving DNELs in the REACH framework (Johansson et al. 2016) and were often not sufficiently accounted for by bodies deriving OELs (Johansson et al. 2012). SCOEL states that variability of workers is taken into account by a variability factor (with toxicokinetic and –dynamic reasons), but does not elaborate on the criteria and size of the factor with regard to genetic disposition, impaired health status or other conditions (SCOEL 2017). In the European Commission's draft guidance on setting AOELs for plant protection products it is concluded that "It is probable that genetics will determine inter-individual variability to the same or a greater extent than age, gender or general health status, therefore the default inter-individual variability factor of 10 is applicable to all exposed groups" (EC 2006).

Areas, where at least based on the methodological descriptions a large degree of convergence exists are

- application of route-to-route extrapolation (and the limitations of this approach)
- the use of substance-specific assessment factors, where data allows
- the principal need to adapt exposure conditions (between experimental animal studies and the exposure situation at workplaces), although detailed guidance on how to perform the adaptation is not always given.

Lack of detailed guidance seems to be a major source of divergent assessments (Maier et al. 2015; Schenk and Johanson 2010). The Joint Task Force of RAC and SCOEL noted that there are differences in the approaches to deal with uncertainties: whereas

SCOEL members preferred dealing with uncertainty as a whole, RAC advocated for a detailed and transparent documentation of individual influencing factors (ECHA/RAC-SCOEL 2017a). This discussion reflects the different histories of expert committees and their wish to retain enough flexibility for case-by-case decisions versus systems with detailed guidance, with the objective to allow many independent assessors to come to similar decisions.

In a recent project on behalf of DG Employment national systems (28 EU member states and 8 non-EU countries) to derive OELs were analysed and compared. Although only some of the national systems have documentation on how OELs are derived and thus are limiting the possibilities for comparing criteria, some conclusions on major sources of differences were drawn (Kalberlah and Bierwisch 2018). The following table summarises these conclusions on reasons for differences between national OELs.

Table 3-2 Reasons for potential differences in national OELs (EU Member States and 8 non-EU countries – Australia, Brazil, Canada, China, India, Japan, South Korea, USA) (table taken from Kalberlah and Bierwisch (2018), shortened)

Criterion	**Reason for potential differences in resulting OEL**
Definition of OEL	Assessment dimensions may be just “health-based”, or may include technical feasibility and/or socio-economic parameters (note that all OEL and OEL(-analogue) systems discussed here are health-based)
Substance definition	OELs may differ for “soluble” vs. “insoluble” compounds for one chemical group of substances (e.g. inorganic cadmium compounds) or may be handled without discriminating solubility with only one OEL for all group members. Similarly, different salts of a metal could be handled as different or identical entity. If similar compounds are all linked to one OEL, there may be different rules, which of the single compounds is regarded representative for the group of compounds
Protection level	“Very sensitive” groups of workers (e.g. due to polymorphism or multiple sensitivity or airway hyper agility) may be protected to a different degree
Adversity	For example, minor sensory irritation or “nuisance” may be regarded as an adverse or non-adverse effect, depending on expert judgement
Minimum data	Some national committees abstain from establishing OELs if only poor data is available, others find it feasible to derive OELs
Indicative or binding character (national level)	For example, in the Netherlands, there are “private” (indicative) and “public” (binding) OELs, which are established by different procedures and therefore may entail different quantitative OELs
Documentation requirements	Thorough documentation usually leads to more transparency and to more systematic analysis of the criteria in assessment and derivation of OELs with potential quantitative consequences due to different completeness of discussion
Induction of re-evaluations (periodically, international	This criterion links to timelines of scientific data and methodological updates, which significantly influences current OELs in place

Criterion	Reason for potential differences in resulting OEL
initiation, case-by-case-new data)	
Basic scenario for workplace exposure assumptions (work-life, working hours/day etc.)	Most OELs refer to 40 years of exposure (full shift, i.e. 8 hrs/day; 5 days/week; 48 weeks per year), however, with few exemptions
Particle fractions (applicable size distribution: inhalable fraction, respirable fraction, total dust)	Adequate transformation from respirable or total dust to inhalable dust scenario may be needed for equivalent protection levels. Similarly, study results with particles (inhalable fraction) may not be used for workplace exposure to fumes with submicron - sized particles without adaptions. These necessary transformations are heterogeneously handled, when OELs are established for the respective particle fractions
Sensory irritants	Different handling of this criterion in time extrapolation and variability; different expert opinions on how to handle animal data on sensory irritation
Selection of relevant study and relevant species, reliability demands	Key criterion for OEL assessments is the “expert opinion” on the quality of data, sometimes guided by consensus quality criteria (e.g. Klimisch Score), but not unambiguously avoiding differences in expert discretion
Default interpretation of assessment factors	Some methodologies are based on a default system of assessment factors or uncertainty factors, others reject any default; moreover, reasons to deviate from defaults may be heterogeneous. Consideration of defaults is often linked to different appreciation of expert opinion, being able to quantify extrapolation factors in case of poor data with or without statistical standard assumptions (i.e. defaults)
Starting point: human data	It is generally assumed that human data are to be preferred to animal data in assessment strategies. However, this has to be weighed against quality of study data (human or animal, respectively) and therefore is a relevant reason for discrepancies in OELs
Intraspecies extrapolation (targeting sensitive individuals) (variability aspects: toxicokinetic/ toxicodynamic)	Workers are often assumed to be a more homogeneous group of exposed persons compared to general population; respective assumption differs. Based on human studies, the minimum size of the exposed group with effect observations as reason to reduce or maintain a default intraspecies factor, is a matter of discussion
Route-to-route extrapolation	Reasons where route-to-route extrapolations are tolerated or not tolerated differ in specific assessments (e.g. because of consideration of “first-pass effects” and potential local effects, and pathway specific absorption)
Safety factors for severity of effects (if any)	In some OEL methodologies certain effects (like reproductive effects or threshold carcinogenic effects may be addressed by including as severity factor)
Safety factors for adequacy of data (if any)	Some OEL methodologies may include a modifying factor in case of poor data (low reliability) or missing data on specific toxicological endpoints

Criterion	Reason for potential differences in resulting OEL
Reproductive toxicants (a) reproductive function	Reproductive toxicants are often differently assessed with regard to adversity (e.g. slight libido effects), time extrapolation (minimum duration of tests to assume qualitative and quantitative coverage of endpoint), interspecies extrapolation (e.g. minor change in sperm counts in rodents may be differently assessed with respect to human relevance) and intraspecies extrapolation (e.g. intraspecies variability from endocrine effects)
Reproductive toxicants (b) developmental effects (if not excluded)	Reproductive effects on the neonates are often not covered in OEL assessments, as only adults are exposed directly at the workplace.
Skin- or airway-sensitising chemical agents	Usually, OELs do not (or only to a limited degree) cover protection from skin- or airway-sensitising effects. However, some OEL systems may partially consider this endpoint, others do not. This discussion also includes adversity assessment, e.g. on preclinical respiratory effects
Non-genotoxic carcinogens	There is high uncertainty on quantitative assessment of a non-stochastic mode of action for carcinogenicity. This is regarded as a key factor for heterogenic OELs for this type of substances
Special rules (e.g. mixtures, UVCB, etc.)	For mixtures national OELs rarely provide unambiguous rules. However, sometimes the additivity rule is mentioned. However, because of different definitions of the “similarity criterion” (where substances are regarded as sufficiently similar that an additivity assumption is regarded as being justified), the practical outcome may be considerably different

The conclusions from this report on differences between national systems for deriving OELs show a high agreement with the critical factors identified and discussed above. Main conclusion also in this report is that there are many areas where the OEL systems would benefit from more guidance and harmonisation. The critical issues identified and reported in the table above can be solved by

- properly defining the values with their objectives and safety goals
- providing guidance on how to perform individual steps in deriving OELs, also with regard to specific endpoints such as reproductive toxicity
- documenting the process for each evaluation.

Recommendations

Detailed guidance should be available for all individual systems deriving OELs or OEL-analogue values. Specifically, guidance should be made available for the following topics:

- Selection of suitable POD (e.g. from several key studies or in a weight-of-evidence approach)

- Modification of the POD to account for exposure conditions of workers

- Step-by-step application of assessment factors and conditions for deviating from defaults

Harmonisation should be pursued for following aspects:

- Harmonisation is required on whether (and if yes under which conditions) LOAELs (together with extrapolation to a NAEL) can serve as a POD

- A harmonised approach (together with detailed guidance) is required for the use of the BMD method for determining the POD

- There is a clear discrepancy with respect to the use of allometric scaling between the PPP and biocides area compared to all other regulatory areas

- Quantification of assessment factors needs harmonisation (e.g. with regard to time and intraspecies extrapolation).

- Necessity and provision of specific assessment factors for specific endpoints, such as reproductive toxicity (comprising of endpoints such as male and female fertility and developmental toxicity).

3.4 Methodological steps for deriving OELs for local respiratory effects

As already explained above, the weight given to local respiratory effects depends on the regulatory area. These kind of effects form the basis of OELs for a large quantity of substances evaluated by OEL expert committees and accordingly takes up a lot of space, e.g. in the SCOEL documentation (SCOEL 2017), but also in the REACH guidance with regard to workers DNELs ((ECHA 2012). However, there is only rudimentary explanations in the biocides guidance (with reference to the REACH guidance) and practically nil in the draft guidance on deriving AOELs for plant protection products. The draft guidance for deriving OELs by RAC is not explicit in this regard. As frequently references are made to the main body of the guidance R.8 as well as to the SCOEL 2017 methodology, it is not easy to decide which document RAC intends to follow.

With regard to other key steps (determining on PODs, assessment factors), as for systemic effects, there exist relevant differences, which might lead to numerical differences in derived values. Few documents comment on the suitability of the benchmark dose approach for respiratory tract toxicity. Quantitative differences between proposed default values exist for all assessment factors.
All documents explicitly addressing evaluation of local effects also make clear that sensory irritation is considered an important endpoint, on which OELs can be based. But differences exist. Whereas under REACH mouse data on respiratory depression (Alarie test) are considered adequate only for short-term DNELs it seems that OEL-deriving expert bodies are basing evaluations also on this kind of data, in case sensory irritation is considered the critical endpoint and adequate human data are lacking. In Germany, although not mentioned explicitly in the guidance papers (BekGS 901 and

TRGS 900), the publication of Brüning et al. (2014) is generally accepted and applied for evaluating local irritating effects (including sensory irritation).

Current approaches provide little information on how to address specificities and potential interspecies differences of particle exposure. In Germany, a proposal was developed in the context of carcinogenic substances in particle form and derivation of exposure-risk relationships (AGS 2013). This proposal includes consideration of species differences in deposition and clearance and the application of deposition models (MPPD). Although not formally imbedded in the guidance documents for deriving OELs relevant elements are also applied for non-carcinogenic substances in particle form for deriving OELs. This approach will further be discussed in a separate report.

Recommendation:

- Guidance is required for selecting suitable PODs for local effects
- Guidance documents should include guidance on evaluating aerosol deposition and clearance in the respiratory tract
- Quantification of assessment factors for local effects (respiratory toxicity, sensory irritation) needs harmonisation.

3.5 Published analyses of existing OELs and similar values

Deveau et al. identified the following issues as possible reasons for OELs showing numerical differences between organisations (Deveau et al. 2015):

- date of derivation of the OEL and timeliness of the database
- selection of key studies (including decisions as to whether use unpublished information)
- different priorities given to human versus experimental data
- selection of PODs, which might differ, e.g. due to varying definitions of adversity,
- lack of harmonisation of application of uncertainty factors
- and the way how additional information is taken into account in a weight-of-evidence approach.

The authors also noted lack of guidance for departing from defaults. These conclusions are supported by the detailed analysis presented in the tables in the chapters above.

A vast range of national and international bodies derive OELs, and differences in methodology and numerical exposure levels also exist between OELs set by national or supranational expert bodies specifically engaged with occupational safety and health issues (Schenk et al. 2008a, b).

Schenk and Johanson analysed existing OELs derived by SCOEL in 2010 (Schenk and Johanson 2010) and again in an update of this work in 2018 (Schenk and Johanson 2018). Both evaluations noticed that a higher margin of safety was used for deriving an OEL when individual factors accounting for uncertainty and variability were

explicitly discussed and numerically defined. The authors also noted that despite discussion on differences between OELs and DNELs uncertainty factors were not more explicitly used by SCOEL in recent years; also, the size of (combined) uncertainty factors did not increase (Schenk and Johanson 2010, 2018).

Considering the methodological differences between OELs as set by SCOEL in past years and REACH DNELS the latter are expected to be systematically lower than OELs. This was confirmed by Schenk and Johanson, when applying the ECHA Guidance on a set of substances and comparing resulting DNELs with existing SCOEL OELs (Schenk and Johanson 2011). Also, Kreider and Williams with the example of styrene concluded that DNELs according to REACH Guidance would be lower than existing OELs (Kreider and Spencer Williams 2010). This was not found by authors when comparing existing DNELs as derived in registration dossiers with OELs (Nies et al. 2013; Schenk et al. 2015; Schenk et al. 2014; Tynkkynen et al. 2015). Obviously, more guidance on when and how to deviate from defaults is required (Schenk and Johanson 2011).

Schenk et al. (2015) also observed a very high variability (up to several orders of magnitude difference) in DNELs from registration dossiers, when comparing it with Swedish OELs and with DNELs for 20 substances derived by the authors themselves, based on the R.8 Guidance document. In many cases the authors observed differences in the dose descriptor used (i.e. differences in selection of key study and leading effect). When comparing DNELs with Swedish OELs then – despite the high variability – on average they were numerically similar. Partly, this can also be explained by the use of existing national or EU OELs as surrogates for DNELs. These authors also noted that for many substances they considered non-thresholded (e.g., genotoxic carcinogens), in registration dossiers a DNEL was derived.

From these studies it can be concluded that deviations might occur if

- no guidance is available
- guidance is insufficient (especially with regard to how and when deviating from defaults is possible or reasonable)
- and/or guidance is not followed.

It should be emphasised that these principles apply primarily to values, which have as a common basis that they are purely health-based. Deveau et al. discussed risk policy decisions such as technical and economic feasibility, which might also be considered by regulatory bodies when setting binding OELs (such as BOELVs) (Deveau et al. 2015). These criteria can differ substantially from country to country. A clear distinction between health-based and other criteria (following the definitions and distinction of risk assessment and risk management) is therefore important to understand the nature of a proposed OEL and to move toward harmonisation.

3.6 General conclusions

This comparison analysed methodological or guidance documents describing how OELs (or other health-based guidance values intended for workers in various regulatory areas) are derived. There is a high accordance in definitions and general

objectives. All systems investigated aim at deriving (purely) health-based values intended to provide protection also for sensitive groups.

But from the analysis presented in the chapters above, as well as from published literature, which comprises publications providing quantitative comparisons of values for the same chemicals and analysing reasons for differences (see section 3.5), it becomes clear that deviations can occur at each individual step of the derivation process:

- data searches and selection of databases for evaluation
- prioritising information (e.g. weighing human versus animal data) and selection of key studies
- concluding on POD(s)
- application of assessment factors (and deviation from defaults)
- adjustment to human (exposure) conditions
- weight-of-evidence considerations of additional information.

Differences were noted not only between approaches used in different regulatory areas. A persistent and unanimous topic in literature is the absence of (sufficient) guidance: Many national committees do not have their detailed methodological approach documented in a guidance document (Kalberlah and Bierwisch 2018; Maier et al. 2015). Lack of documentation on methodologies used for deriving OELs is generally seen as the reason for lack of transparency and a hindrance for harmonisation. Therefore, complete and detailed documentation of methods used is desirable, which should not only describe the default approach, but also when and how deviation from defaults is possible. In order to achieve the overall goal of transparency and harmonisation, the following elements are necessary

- availability of detailed guidance
- compliance with guidance
- transparent documentation of evaluations.

Annex 1

The tables in this Annex document details retrieved from the respective method documentation (and in some cases complemented by personal communication). It can be used for reference and to trace overall conclusions drawn. The wording in these tables is kept close to the wording in the respective documents. A difference in wording does not necessarily implicate a real difference in the approaches.

Lines, which further analyse specific issues by differentiated questions following the initial question, are shaded grey.

Example:

Modification of dose descriptor?
Correction for differences in route-specific absorption between animal and humans
Correction for different exposure conditions
Correction for different physical activity

Analysis of scope and definitions of OELs and OEL-analogue values

Table A-1 Comparison of scope and definition of OEL values

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**EU PPP Directive**	**EU Biocidal Products Regulation**
Name	DNEL (derived no effect level)	OEL (occupational exposure limit)	OEL (occupational exposure limit)	AGW (Arbeitsplatz-grenzwert)	MAK (maximum workplace con centration)	DNEL (derived no effect level)	AOEL (acceptable operator exposure level)	AEL (acceptable exposure level)
Target population	General public & workers	Workers	Workers	Workers	Workers	General public & workers	Operators (applying plant protection products) and workers, but should also be applicable to bystanders and residents	Professionally and non-professionally exposed humans
Unit(s)	mg/m³ (or ppm)	mg/m³	mg/m³ (or ppm)	mg/m³ (or ppm)	ppm or mg/m³	Not addressed, but can be assumed to be in line with ECHA Guidance R.8	mg/kg bw/day	mg/kg bw/day
Specification of values	Long-term: daily dose	8-Hour time weighted	8-Hour time weighted	TWA for work life exposure	8-Hour time weighted	8-Hour time weighted	Internal (absorbed)	Internal (absorbed)

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**EU PPP Directive**	**EU Biocidal Products Regulation**
	(workers: 8 h per day, 40 years, 48 weeks/year; 5 days/week)	average (TWA), 40 years (48 weeks/year; 5 days/week; i.e. 9600 days or 76,800 hours);	average (TWA) , 40 years (48 weeks/year; 5 days/week; i.e. 9600 days or 76,800 hours)	at 8 hours, 5 d/w	average (TWA), 40 h/week, working life (duration of working life not specified)	average (TWA)	dose available for systemic distribution from any route of exposure	dose available for systemic distribution from any route of exposure, expressed as internal levels; should follow same common scientific principles as AOELs for PPP (refe-rence to EC guidance doc.)
Protection goal	'Overall' No-Effect-Level for a given exposure (…), accounting for uncertainties/v ariability in these data and the human population ex-posed, Part B; level of exposure	OELs established to protect workers from adverse effects on health	Levels of exposure below which no detrimental effects are expected	Protection against adverse health effects following occupational chronic and acute inhalation exposure; concentration at which in general no	Maximum concentration which generally does not have known adverse effects on the health nor cause unreasonable annoyance	Not specified	The maximum amount of active substance to which the operator may be exposed without any adverse health effects	With reference to the AOEL definition: "... the maximum amount of active substance to which the operator may be exposed without any adverse health effects"

	REACH Regulation	RAC OEL metho-dology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	EU PPP Directive	EU Biocidal Products Regulation
	above which humans should not be exposed; below DNEL risk to humans can be con-sidered to be controlled, R.8			adverse health effects are expected				
Includes susceptible groups?	Yes	Yes	Yes	Yes	Yes	Yes	Implicit, application of variability factor	Yes, all (no differentiation between professionals and non-prof.)
Does include developmen-tal toxicity and intends to be protec-tive for wo-men in child-bearing age?	Yes	Yes	Pregnant women in scope in new guidance, but existing evaluations in conflict, see section 3.1	No; notation to provide information whether developmental toxicity at OEL is excluded	No; notation to provide information whether developmental toxicity at OEL is excluded	Yes	Yes	Yes
Other endpoints such as endocrine	All available data (endocrine effects mentioned as	Not mentioned	Not mentioned	Not mentioned	Not mentioned, but in practice considered*	Not mentioned	Whole toxicological database (e.g.	All toxic effects, including

	REACH Regulation	RAC OEL metho-dology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	EU PPP Directive	EU Biocidal Products Regulation
disruptive activity?	possibly mechanism for reproductive toxicity)						…endocrine effects)	endocrine disruption, …
Does consider respiratory sensitisation ?	Only qualitative assessment possible	Respiratory sensitisation notation assigned; dose-response data to be used for OEL, if data allow	Respiratory sensitisation notation assigned; to be used for OEL, if data allow	Respiratory sensitisation notation assigned	Respiratory sensitisation notation assigned	Not specifically mentioned	Not specifically mentioned	Qualitative approach preferred for this endpoint
Does consider sensory irritation?	Yes	Yes	Yes	Yes	Yes	Yes	Not specifically mentioned	Not specifically mentioned
STEL values derived?	DNELacute for short periods (from minutes to a few hours).	Yes, STEL 15 min; refers to SCOEL, 2017	Yes, STEL 15 min; needed when a relevant effect is observed following a brief exposure	Yes, STEL 15 min; excursion factors depending on local or systemic action	Yes, STEL 15 min; excursion factors depending on local or systemic action	$DNEL_{acute}$ for short periods	Acute AOELs, some values were derived during sub-stance appro-val or renewal evaluation, but agreed guidan-ce pending**	Acute, medium- and long-term AELs, for professionals focus on medium- and long-term AELs

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**EU PPP Directive**	**EU Biocidal Products Regulation**
How is the dermal exposure route considered	quantitative DNELs for dermal route to be derived	assignment of skin notations	assignment of skin notations	assignment of skin notations	assignment of skin notations, derived by dermal uptake in relation to the systemic NOAEL	quantitative DNELs for dermal route to be derived	(systemic) AOEL to be used for assessing dermal route	(systemic) AEL to be used for assessing dermal route

* Personal communication, MAK Commission, December 2020
** COMMISSION GUIDANCE DOCUMENT, SANTE-10832-2015 rev. 1.7, 24 January 2017; Guidance on the assessment of exposure of operators, workers, residents and bystanders in risk assessment for plant protection products; https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_ppp_app-proc_guide_tox_accpt-exp-levs-2015.pdf

Analysis of requirements for data searches and evaluation

Table A-2 Comparison of requirements on data searches and evaluation and documentation of information used for assessments

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
Requirements for information searches and type of information to be used	Any relevant hazard information that is available, including human, animal, in vitro, in silico data	Published reviews and assessments, REACH registration dossiers and peer reviewed literature	All available data; Systematic information search descri-bed; call for submission of data	Not mentioned	Published studies and database entries. If necessary, also unpublished studies from industry, if full text is on hand	Not in scope	All studies required under Directive 91/414/EEC	Collect all available information on toxicological properties including animal, in vitro, in silico and human data
Explicit data source mentioned?	no (but ECHA Guidance on IR&CSA, R.3 is implicitly applicable)	no	Yes; systematic and detailed description of data search procedure given	No	No	Not in scope	No	Yes
If yes, which data source is mentioned?	-	-	Databases e.g. PubMed, SCOPUS,	-	-	-	-	eChemPortal and QSAR Toolbox

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
			ScienceDirect, SciFinder					Additional list of sources: refers to ECHA Guidance on IR&CSA, R.3
Assessment of data quality	Data to be assessed for reliability and consistency … taking into account the quality of the testing method, size and power of the study design, biological plausibility, dose-response relationships and statistical association	Assessment of adequacy, relevance and reliability refers to ECHA Guidance on IR&CSA, R.4, and R.8-15 and to SCOEL 2017 mentions tools for systematic reviews for quality assessment	Adequacy of the evidence for each hazardous property is evaluated	Not mentioned	Case by case decision whether a study is relevant (in practice similar to REACH, for human studies refers to DFG (1997) for details)*	Assessment of reliability and consistency. Shortcomings entail higher AF for quality of the database	Setting of a “provisional” AOEL if data quality departs significantly from standards until further data is submitted	Assessment of completeness, reliability and relevance. Use of scoring criteria (e.g. Klimisch) to weigh relevance of data; refers to ECHA Guidance on IR&CSA, R.4
Identification of critical	Expert judgement	Most relevant adverse	Consistency of all available	Selection of most relevant	Most sensitive endpoint with	Consideration of type of	Most sensitive relevant	Biological plausibility of

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
effects / key studies		effect(s), leading to the most appropriate limit value; iterative process	evidence considered	study, inhalation studies preferred, most sensitive species preferred.	human health relevance	human exposure. Chosen NOAEL not necessarily the lowest, but the most relevant for humans. More than one critical effect and key study are possible. Human data in good quality should be used when available	endpoint. However, if a substance has severe effect with a higher NOAEL, it should be departed from the severe effect with an additional AF, reflecting an increased margin of safety	dose-effect relationship, severity and reversibility of effects, mode of action, human relevance
Weight-of-evidence (WoE) elements	Yes, evaluation of the entire body of available data for consistency and biological plausibility. High quality	Yes, assessment of the relative weights of different pieces of the available information, incl. MoA,	Yes, see above	Yes, if several repeated-dose studies: carefully select most relevant study	Not mentioned, in practice similar to REACH, RAC or AGS*	Yes	Yes, whole dataset and all type of effects need to be considered, mechanistic aspects as well as relevance of	Yes, consistency over dataset to be considered

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
	given more weight than lower quality.	quality of the data, consistency of results, nature and severity of effects, and relevance					effects in humans	
Application of read-across and QSAR	Yes, detailed guidance available	Yes, non-testing data (e.g. read-across) to be considered	Not system-atically, but in some cases data from related substances used as support	Not mentioned	Read-across via expert judgement on case-by-case basis*	Reference to other substances only to support quantification of assessment factors	Not mentioned	Yes, refers to REACH guidance
Use of human data	Yes, human data are an appropriate basis also for the derivation of DNEL	Yes, human data of good quality are given preference	Yes. generally, human studies with populations encompassing workers are more relevant than animal studies	Yes	Yes, human data is principally given the highest weight. longitudinal studies usually necessary	Yes, refers to TR 104 regarding details	Used as supplementary information to confirm the validity of AOEL, which is based on exper. data; if appropriate scientifically valid and	Yes “Relevant human data of an adequate quality can sometimes be the best available data but, more frequently, the available

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
							ethically generated human data are available and lead to lower AOEL, they should be used	human, animal, and other data are considered together in order to reach a conclusion
Consideration of severity of effects?	Considered as part of dose-response considerations	Part of WoE consideration	Not mentioned	Not mentioned	Addressed only in context of sensory irritation: Higher AF for severe effects at the LOAEC when extrapolating to NAEC, in practice considered on a case by case basis*	Higher AF for LOAEL-NAEL extrapolation in case of severe effects	Yes, expert judgement	For severe reproductive effects, the need for an additional assessment factor should be considered
Update of evaluation	Re-evaluation if further hazard information	Not mentioned	OELs need to be frequently reviewed as science	Not mentioned	Not mentioned in methodolo-gical des-cription;	Not in scope	Not mentioned	Not mentioned

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
	becomes available		progresses, new evidence becomes available and experience is gained		”announce-ment list” in OEL list indicate substances in reevaluation			
Documenta-tion require-ments for data and assess-ments	Part of regulatory process, detailed derivation not published	Not mentioned (no information on content of opinions)	Not mentioned (no information on content of opinions)	Not mentioned (no information on content of opinions)	Detailed scientific justifications published (no further details)	Not in scope	Part of regulatory process	Part of regulatory process

* Personal communication, MAK Commission, December 2020

Analysis of steps for deriving OEL or OEL-analogue values for systemic effects

Table A-3 Comparison of methodologies to derive a systemic OEL

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
Minimum data requirement for deriving value	Annex VII information requirements allow deriving DNEL	Not mentioned	Not mentioned	Not mentioned	Sufficient toxicological or human data required (not further specified)	Not mentioned (data according to REACH information requirements assumed)	Not specifically mentioned, based on data according to requirements in Directive 91/414/EEC	Not specifically mentioned, based on complete dataset according to information requirements
Applicable POD or dose descriptors	NOAEL, LOAEL, BMD(L)	E.g. BMD(L) or NOAEL	NOAEL, BMD(L);	NOAEL, BMD(L)	Mentioned only for irritation (see below), although NOAEL, BMDL and LOAEL applied in practice also for systemic effects*	NOAEL, implied LOAEL, BMD(L)	NOAEL, BMD(L), (LOAEL only, if NOAEL or BMD not possible)	NOAEL, LOAEL, BMD(L)

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
If BMD(L), specification whether BMD or BMDL and critical effect size for quantal data?	BMDL05 (if other dose descriptors, then use additional factor according to expert judgement)	BMD, not further specified (for non-carcinogenic effects)	BMD05	not specified	$BMDL_{05}$ or $BMDL_{SD}$*	BMDL, Critical effect size: if 5% is used, no extrapolation factor to NAEL required	Not specified (presumably BMDL)	Not specified
Selection of POD	POD resulting in lowest DNEL	POD resulting in lowest limit value	Most relevant effect, based on the entire available database, considering consistency and interdependence of effects and mechanisms	POD from most relevant study, performed with most sensitive species	NOAEL of most sensitive endpoint with human health relevance	Not in scope, refers to R.8; refers to TR 104 regarding POD from human data	Lowest NOAEL in sensitive species, but if severe effect, requiring high assessment factor leads to lower AOEL, this effect should be used	The one resulting in the lowest AEL
Route-to-route extrapolation applicable?	Yes, although not recom-mended; applicability	Not discussed	Yes, although not recom-mended. Refers to R.8	Yes, specific criteria listed	Yes	Yes, but advocates against default factors	Yes, often necessary, as toxicological data are often via the oral	Yes, but consider potential route-specificity

	REACH Regulation	RAC OEL metho-dology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	Plant Protection Products Directive	EU Biocidal Products Regulation
	has to be checked						route; provided there are no indications for route-specificity	
Modification of POD?	Yes , detailed guidance given	Yes, but not explained in detail	Yes, details given	Yes, but not explained in detail	Yes	Yes, but not in scope; refers to R.8	Yes, e.g. for conversion of concentrations in feed and drinking water	Yes, general reference to ECHA Gui-dance on IR and CSA, R.8
Correction for differences in route-specific absorption between animal and humans	Yes	Not mentioned	Yes	Yes	Yes	Yes; refers to R.8	Yes	Yes
Correction for different exposure conditions (e.g. exposure duration)	Yes	Refers to SCOEL methodology 2017	Yes	Not mentioned	Yes	Yes; refers to R.8	Yes (oral pathway)	Yes, including different time patterns of exposure

	REACH Regulation	RAC OEL metho-dology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	Plant Protection Products Directive	EU Biocidal Products Regulation
Correction for different physical activity	Yes	Refers to SCOEL methodology 2017	Yes	Not mentioned in resp. guidance, but in practice AGS (2013) is followed; assumes 10m³/8 h for slight physical activity	Yes	Yes; refers to R.8	No	Yes
Default anthropo-metric data provided?	Yes	Not specifically mentioned in R.8.17	No. implicitly uses R.8 values	No	Yes	Refers to R.8	Yes (for oral pathway)	No
Use of assessment factors (AF)?	Yes	Yes	Yes, partly	Yes	Yes, partly	Yes	Yes	Yes
Chemical-specific assessment factors?	Yes, if available	Yes, if available	Yes, if available	Yes, if available	Yes, if available	Yes, if available	Yes, if available	Yes, if available
Default AF for time extrapolation	sa** – c: 6 sa – sc: 3 sc – c: 2	Yes; refers to R.8.4.3: sa** – c: 6	No	sa** – c: 6 sa – sc: 2 sc – c: 2	Mentioned only for irritation (see below),	sa** – c: 6 sa – sc: 3 sc – c: 2	sa** – c: - sa – sc: - sc – c: 2	sa** – c: 6 sa – sc: 3 sc – c: 2

	REACH Regulation	RAC OEL metho-dology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	Plant Protection Products Directive	EU Biocidal Products Regulation
		sa – sc: 3 sc – c: 2			although factors are applied in practice*			
Allometric scaling for interspecies extrapolation ?	Yes, exponent 0.75	Yes; refers to R.8.4.3	Yes, exponent 0.75	Yes, exponent 0.75	Yes; refers to R.8	Yes, exponent 0.75	No	No, but can be used to replace default AF
Default AF for interspecies extrapolation	2.5	Yes; refers to R.8.4.3	No default provided	Inter + Intra = 5	Inter + Intra = 2*	1	10	10
Default AF for intraspecies extrapolation	Worker: 5	Yes; refers to R.8.4.3	>=1	Inter + Intra = 5	Inter + Intra = 2*	Worker: 3	10	10
AF for severity of effects	1, should be increased on a case-by-case basis	Yes; refers to R.8.4.3	Not explicitly addressed	Not considered necessary	Only mentioned for sensory irritation; in practice considered on a case by case basis*	1, requires larger AF if severe effects at LOAEL	≤ 10 on a case-by-case basis, (e.g. for teratogenic or irreversible neuropathic effects)	2 -10, taking into account dose-response data

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
AF for LOAEL-NAEL extrapolation	Yes; 3 – 10	Yes; refers to R.8.4.3	Use NOAEL or BMD	Use NOAEL or BMDL	Only mentioned for sensory effects, (2-3); in practice also applied for systemic effects*	3; BMDL05 is considered NOAEL	Only, if BMD not possible, no default, case-by-case basis	Yes, no default value, depends on slope, last resort
Quality of database	1, should be increased on a case-by-case basis	Yes; refers to R.8.4.3	Yes, no default provided	Not mentioned	Not mentioned, but considered by expert judgement*	1, should be increased on a case-by-case basis; referen-ces ECETOC TR 104	Yes, no default provided	Yes, no default provided
Use of PBPK modelling, where applicable?	Yes	Not discussed	Yes	Yes, substance-specific refinement	Yes	Yes, may be used to refine AF	Yes, if sufficient information available	Yes, possible as a refine-ment step, reference to WHO/IPCS document
How is uncertainty and variability addressed?	Assessment factors are applied to address	Not discussed in detail, application of AF	Overall uncertainty factor to account for	Not discussed in detail, application of AF	Not discussed, discussion of uncertainties and variabilities in	Assessment factors account for both	Application of additional AF according to expert judgement	For local effects reference to Guidance on

	REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
	uncertainty and variability.		uncertainty and variability; rejection of OEL reco-mmendation in case of inade-quate data		MAK value documentation	uncertainty and variability		IR and CSA R.19

* Personal communication, MAK Commission, December 2020

** sa: subacute, sc: subchronic, c: chronic

Analysis of steps for deriving OEL or OEL-analogue values for local effects

Table A-4 Comparison of methodologies to derive a local inhalation OEL

	REACH Regulation	**RAC OEL methodology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
Applicable POD or dose descriptors	NOAEC, LOAEC (no specific mentioning of BMD in the context of local effects)	No specific provisions for local effects	Clear distinction between local and systemic effects, but no specific provisions with respect to the following	NOAEC, BMD	NOAEC, BMDL or LOAEL, mentioned only for sensory irritation; however, in practice also used for local effects*	No specific provisions for local effects	See footnote ***	NOAEC
Modification of POD?	Yes	-	Yes	No explicit explanation for local effects given in TRGS 900, but guidance for carcinogens in TRGS 910 (AGS 2013) used in practice	Yes	See below	-	No explicit explanation for local effects given

	REACH Regulation	RAC OEL methodology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	Plant Protection Products Directive	EU Biocidal Products Regulation
Correction for different exposure conditions	Yes, but not for effects that are driven by concentration (e.g. irritation)	-	No explicit explanation for local effects given	Yes, see AGS (2013)	Yes	Not for concentration-driven effects	-	No explicit explanation for local effects given
Correction for different physical activity	Only if there is evidence that not concentration-driven	-	No explicit explanation for local effects given	Yes, see AGS (2013)	Yes, but not for concentrat-ion-driven effects	Not for concentration-driven effects	-	No explicit explanation for local effects given
Use of assessment factors (AF)?	Yes	General reference to R.8	For sensory irritation	Yes	For sensory irritation, however, in practice also used for local effects*	Yes	-	Yes
Default AF for time extrapolation	sa** – c: 6 sa – sc: 3 sc – c: 2 (lower, if effects are concentration-dependent)	-	No explicit explanation for local effects given	sa** – c: 6 sc – c: 2	sa** – c: 6 sc – c: 2 for sensory irritation; however, in practice also used for local effects*	1 for all time extrapolations	-	Reference to REACH assessment factors is made

	REACH Regulation	RAC OEL methodology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	Plant Protection Products Directive	EU Biocidal Products Regulation
Allometric scaling for interspecies extrapolation?	No – assessment based on concentration	-	No – assessment based on concentration	No – assessment based on concentration	No – assessment based on concentration	No – assessment based on concentration	-	No
Default AF for interspecies extrapolation	2.5	-	Specific correlation for sensor. irritation (Alarie) in annex	Inter + Intra = 5	inter + intra = 2, for sensory irritation; however, in practice also used for local effects*	1	-	2.5 (based on air concen-tration)
Default AF for intraspecies extrapolation	Worker: 5	-	2 for sensor. irritation, no default for other local effects	Inter + Intra = 5	inter + intra = 2, for sensory irritation; however, in practice also used for local effects*	Worker: 3	-	10 (for professionals and non-professionals)
AF for severity of effects	1, should be increased on a case-by-case basis	-	No explicit explanation for local effects given	Not mentioned	Not explicitly addressed, considered on a case by case basis*	No specific provisions for local effects	-	Reference to REACH assessment factors

	REACH Regulation	RAC OEL methodology	SCOEL	AGS – German OELs	DFG MAK	ECETOC	Plant Protection Products Directive	EU Biocidal Products Regulation
AF for LOAEL-NAEL extrapolation	Yes, 3 - 10	-	2 – 3 for sensor. irritation, no default for other local effects	LOAEL not foreseen as POD, although in practice sometimes used with factor 3	2 - 3	No specific provisions for local effects	-	Reference to REACH assessment factors
Quality of database	1, should be increased on a case-by-case basis	-	No explicit explanation for local effects given	Not mentioned	Not mentioned, in practice considered by expert judgment*	No specific provisions for local effects	-	Reference to REACH assessment factors
How is sensory irritation considered?	Important endpoint for acute and chronic workers DNELs (animal data – Alarie test - to be used for short-term values only)	Not mentioned	Important endpoint for deriving acute and chronic OELs; specific annex on this subject	Not specifically mentioned in BekGS 901, but Brüning et al. (2014) accepted as guidance	Sensory irritation (human data preferred, but also Alarie test) is taken into account when establi-shing OELs, Brüning et al. accepted as guidance	Sensory irritation is discussed with a focus on adversity of these effects	-	Not explicitly mentioned
How is deposition and clearance	Differences in deposition of particles and	Not mentioned	Deposition of particles in the airways is given	Methodological approach for carcinogens also	Not mentioned,	Deposition is addressed (in TR 86) and	Not mentioned	Differences in deposition of particles and

	REACH Regulation	**RAC OEL methodology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**Plant Protection Products Directive**	**EU Biocidal Products Regulation**
of aerosols in the respiratory tract considered?	fate in respiratory tract to be covered by interspecies factor		as an example for potential differences between humans and experimental animals, which is given consideration; no further details	used in practice for non-carcinogenic endpoints, see AGS (2013)	in practice HEC calcu-lation (AGS 2013) is used*	models (MPPD) are recommended for particles		fate in respiratory tract to be covered by interspecies factor

* Personal communication, MAK Commission, December 2020

** sa: subacute, sc: subchronic, c: chronic

*** if local effects occur in inhalation studies, systemic AOEL should be adequately protective for local effects; if local effects predominate, OEL-analogue value can be derived; this is not further detailed in EC (2006) or related documents

References

AGS, Ausschuss für Gefahrstoffe. (2010). Bekanntmachung zu Gefahrstoffen. Kriterien zur Ableitung von Arbeitsplatzgrenzwerten. BekGS 901. Das Gemeinsame Ministerialblatt 61 (32), 691-696. http://www.baua.de/cln_135/de/Themen-von-A-Z/Gefahrstoffe/TRGS/Bekanntmachung-901.html.

AGS, Ausschuss für Gefahrstoffe. (2013). Leitfaden zur Quantifizierung stoffspezifischer Expositions-Risiko-Beziehungen und von Risikokonzentrationen bei Exposition gegenüber krebserzeugenden Gefahrstoffen am Arbeitsplatz, (Anlage 3 zu TRGS 910). Version N10, Stand: 15.09.2013, Ausarbeitung durch den Arbeitskreis Risikoableitung im UA III des AGS. http://www.baua.de/de/Themen-von-A-Z/Gefahrstoffe/TRGS/TRGS-910.html.

AGS, Ausschuss für Gefahrstoffe. (2015). Begründung zu Vanadiumverbindungen in TRGS 900. Ausgabe: November 2015. Stand: Mai 2015, online: http://www.baua.de/de/Themen-von-A-Z/Gefahrstoffe/TRGS/Arbeitsplatzgrenzwerte.html__nnn=true.

AGS, Ausschuss für Gefahrstoffe. (2018). Technische Regeln für Gefahrstoffe – Arbeitsplatzgrenzwerte (TRGS 900). Ausgabe: Januar 2006. BArBl Heft 1/2006 S. 41-55. Geändert und ergänzt: GMBl 2018 S.542-545[Nr.28] (v.07.06.2018), Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. https://www.baua.de/DE/Angebote/Rechtstexte-und-Technische-Regeln/Regelwerk/TRGS/pdf/TRGS-900.pdf?__blob=publicationFile&v=11.

Brüning, T., Bartsch, R., Bolt, H.M., Desel, H., Drexler, H., Gundert-Remy, U., Hartwig, A., Jäckh, R., Leibold, E., Pallapies, D., Rettenmeier, A.W., Schlüter, G., Stropp, G., Sucker, K., Triebig, G., Westphal, G., van Thriel, C. (2014). Sensory irritation as a basis for setting occupational exposure limits. Archives of Toxicology 88, 1855-1879.

Deveau, M., Chen, C.P., Johanson, G., Krewski, D., Maier, A., Niven, K.J., Ripple, S., Schulte, P.A., Silk, J., Urbanus, J.H., Zalk, D.M., Niemeier, R.W. (2015). The global landscape of occupational exposure limits—implementation of harmonization principles to guide limit selection. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S127-144.

DFG, Deutsche Forschungsgemeinschaft. (1997). Verhaltenstoxikologie und MAK-Grenzwertfestlegungen, WILEY-VCH Verlag GmbH Weinheim.

DFG, Deutsche Forschungsgemeinschaft. (2018). MAK- und BAT-Werte-Liste 2018. Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe. Mitteilung 54, WILEY-VCH Verlag GmbH, Weinheim. https://onlinelibrary.wiley.com/doi/book/10.1002/9783527818402.

DFG, Deutsche Forschungsgemeinschaft. (2019). MAK- und BAT-Werte-Liste 2019. Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe. Mitteilung 55,

WILEY-VCH Verlag GmbH, Weinheim. https://onlinelibrary.wiley.com/doi/book/10.1002/9783527826155.

EC, European Commission. (2006). Guidance for the Setting and Application of Acceptable Operator Exposure Levels (AOELs). Draft. SANCO 7531 - rev.10, Directorate E – Safety of the food chain. E3 - Chemicals, Contaminants, Pesticides.

ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals. (2003). Technical Report No. 86. Derivation of Assessment Factors for Human Health Risk Assessment, Brussels, Belgium.

ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals. (2010). Technical Report No. 110. Guidance on Assessment Factors to Derive a DNEL, Brussels, Belgium.

ECHA, European Chemicals Agency. (2011). Guidance on Information Requirements and Chemical Safety Assessment. Part B: Hazard Assessment. Version 2.1. December 2011, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

ECHA, European Chemicals Agency. (2012). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1, November 2012, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

ECHA, European Chemicals Agency. (2017). Guidance on the Biocidal Products Regulation. Volume III Human Health - Assessment & Evaluation.(Parts B+C). Version 4.0. December 2017, https://echa.europa.eu/documents/10162/23036412/biocides_guidance_human_health_ra_iii_part_bc_en.pdf/30d53d7d-9723-7db4-357a-ca68739f5094.

ECHA, European Chemicals Agency. (2019). Guidance on Information Requirements and Chemical Safety Assessment. Appendix to Chapter R.8: Guidance for preparing a scientific report for health-based exposure limits at the workplace. Version 1.0 August 2019, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment
ECHA/RAC-SCOEL. (2017a). ECHA/RAC-SCOEL Joint Task Force Report. Scientific aspects and methodologies related to the exposure of chemicals at the workplace. Final Version – 28 February 2017, https://echa.europa.eu/documents/10162/13579/rac_joint_scoel_opinion_en.pdf/58265b74-7177-caf7-2937-c7c520768216.

ECHA/RAC-SCOEL. (2017b). ECHA/RAC-SCOEL Joint Task Force Report. Scientific aspects and methodologies related to the exposure of chemicals at the workplace. TASK 2. 6 December 2017. Final report, https://echa.europa.eu/documents/10162/13579/jtf_opinion_task_2_en.pdf/db8a9a3a-4aa7-601b-bb53-81a5eef93145.

EFSA, European Food Safety Authority. (2006). Opinion of the Scientific Panel on Plant health, Plant protection products and their Residues on a request from the Commission on the Guidance Document (GD) for the establishment of acceptable operator exposure levels (AOELs). The EFSA Journal 4(4):345, 1-12.

EFSA, European Food Safety Authority. (2012). Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data. The EFSA Journal 10(3):2579, https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2012.2579.

Hoffmann, S., de Vries, R.B.M., Stephens, M.L., Beck, N.B., Dirven, H.A.A.M., Fowle, J.R., Goodman, J.E., Hartung, T., Kimber, I., Lalu, M.M., Thayer, K., Whaley, P., Wikoff, D., Tsaioun, K. (2017). A primer on systematic reviews in toxicology. Archives of Toxicology 91, 2551-2575.

Johansson, M.K., Johanson, G., Öberg, M. (2012). How are asthmatics included in the derivation of guideline values for emergency planning and response? Regulatory Toxicology and Pharmacology 63, 461-470.

Johansson, M.K., Johanson, G., Öberg, M., Schenk, L. (2016). Does industry take the susceptible subpopulation of asthmatic individuals into consideration when setting derived no-effect levels? Journal of Applied Toxicology 36, 1379-1391.

Ju-Kun, S., Yuan, D.-B., Rao, H.-F., Chen, T.-F., Luan, B.-S., Xu, X.-M., Jiang, F.-N., Zhong, W.-D., Zhu, J.-G. (2016). Association between Cd exposure and risk of prostate cancer: a PRISMA-compliant systematic review and meta-analysis. Medicine 95, e2708.

Kalberlah, F., Bierwisch, A. (2018). OEL/STEL deriving systems. Project report for: Third study on collecting most recent information for a certain number of substances with the view to analyse the health, socio-economic and environmental impacts in connection with possible amendments of Directive 2004/37/EC. (Ref: VC/2017/0011). Final Report. February 2018, Forschungs- und Beratungsinstitut Gefahrstoffe GmbH (FoBiG), Freiburg, Germany in cooperation with RPA, COWI, and EPRD; prepared for DG Employment, Social Affairs & Inclusion.

Klimisch, H.J., Andreae, M., Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology 25, 1-5.

Kreider, M.L., Spencer Williams, E. (2010). Interpreting REACH guidance in the determination of the derived no effect level (DNEL). Regulatory Toxicology and Pharmacology 58, 323-329.

Lavelle, K.S., Schnatter, R.A., Travis, K.Z., Swaen, G.M., Pallapies, D., Money, C., Priem, P., Vrijhof, H. (2012). Framework for integrating human and animal data in chemical risk assessment. Regulatory Toxicology and Pharmacology 62, 302-312.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 6, e1000100.

Maier, A., Lentz, T.J., MacMahon, K.L., McKernan, L.T., Whittaker, C., Schulte, P.A. (2015). State-of-the-science: the evolution of occupational exposure limit derivation and application. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S4-S6.

Nies, E., Musanke, U., Püringer, J., Rühl, R., Arnone, M. (2013). DNELs for workplaces - observations from an inspection of the DGUV DNEL list. Gefahrstoffe - Reinhaltung der Luft 73, 455-462.

RAC, Committee for Risk Assessment. (2018a). Opinion on scientific evaluation of occupational exposure limits for Acrylonitrile. ECHA/RAC/ O-0000001412-86-188/F. Adopted 9 March 2018, European Chemicals Agency (ECHA), Helsinki, Finland.

RAC, Committee for Risk Assessment. (2018b). Opinion on scientific evaluation of occupational exposure limits for benzene. ECHA/RAC/ O-000000-1412-86-187/F. Adopted 9 March 2018, European Chemicals Agency (ECHA), Helsinki, Finland. https://echa.europa.eu/documents/10162/13641/benzene_opinion_en.pdf/4fec9aac-9ed5-2aae-7b70-5226705358c7.

RAC, Committee for Risk Assessment. (2018c). Opinion on scientific evaluation of occupational exposure limits for nickel and its compounds. ECHA/RAC/ A77-O-0000001412-86-189/F. Adopted 9 March 2018, European Chemicals Agency (ECHA), Helsinki, Finland. https://echa.europa.eu/documents/10162/13641/nickel_opinion_en.pdf/9e050da5-b45c-c8e5-9e5e-a1a2ce908335.

RAC/SCOEL. (2016). Joint Opinion to resolve differences in scientific opinion as regards exposure levels for N-Methyl-2-Pyrrolidone. 30 November 2016, ECHA Committee for Risk Assessment (RAC) and Scientific Committee on Occupational Exposure Limits (SCOEL). https://echa.europa.eu/documents/10162/13579/rac_joint_nmp_opinion_en.pdf/e4b4f43b-a3bd-a7c0-08be-16c3886593e7.

Schenk, L. (2010). Comparison of data used for setting occupational exposure limits. International Journal of Occupational and Environmental Health 16, 249-262.

Schenk, L., Deng, U., Johanson, G. (2015). Derived No-effect Levels (DNELs) under the European Chemicals Regulation REACH - An analysis of long-term inhalation worker-DNELs presented by industry. Annals of Occupational Hygiene 59, 416-438.

Schenk, L., Hansson, S.O., Rudén, C., Gilek, M. (2008a). Are occupational exposure limits becoming more alike within the European Union? Journal of Applied Toxicology 28, 858-866.

Schenk, L., Hansson, S.O., Rudén, C., Gilek, M. (2008b). Occupational exposure limits: A comparative study. Regulatory Toxicology and Pharmacology 50, 261-270.

Schenk, L., Johanson, G. (2010). Use of Uncertainty Factors by the SCOEL in their derivation of health-based Occupational Exposure Limits. Critical Reviews in Toxicology 40, 791-798.

Schenk, L., Johanson, G. (2011). A quantitative comparison of the safety margins in the European indicative occupational exposure limits and the derived no-effect levels for workers under REACH. Toxicological Sciences 121, 408-416.

Schenk, L., Johanson, G. (2018). Use on uncertainty factors by the European Commission Scientific Committee on Occupational Exposure Limits: a follow-up. Critical Reviews in Toxicology 48, 513-521.

Schenk, L., Johanson, G. (2019). Will worker DNELs derived under the European REACH regulation extend the landscape of occupational exposure guidance values? Archives of Toxicology 93, 1187-1200.

Schenk, L., Palmen, N., Theodori, D. (2014). Evaluation of worker inhalation DNELs. Part A: Quality assessment of a selection of DNELs. Part B: Discussion paper on the possibilities to improve the overall quality of DN(M)ELs. RIVM Letter report 110001001/2014, RIVM, National Institute for Public Health and the Environment.

Schuhmacher-Wolz, U., Hassauer, M., Kalberlah, F. (2006). Ableitung von Arbeitsplatzrichtwerten für Stoffe mit reproduktionstoxischen Eigenschaften. Fb 1079, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund/Berlin/Dresden, Wirtschaftsverlag NW Bremerhaven.

Schulte, P., Howard, J. (2011). Genetic susceptibility and the setting of occupational health standards. Annual Review of Public Health 32, 149-159.

SCOEL, Scientific Committee for Occupational Exposure Limits. (2013). Methodology for the Derivation of Occupational Exposure Limits. Key Documentation (version 7), June 2013, European Commission, Employment, Social Affairs and Inclusion.

SCOEL, Scientific Committee for Occupational Exposure Limits. (2017). Methodology for derivation of occupational exposure limits of chemical agents. The General Decision-Making Framework of the Scientific Committee on Occupational Exposure Limits (SCOEL) 2017, European Commission. Directorate-General for Employment, Social Affairs and Inclusion. Directorate B - Employment. Unit B.3 - Health and safety, Brussels. https://publications.europa.eu/en/publication-detail/-/publication/3c8ef3e0-48fc-11e8-be1d-01aa75ed71a1/language-en.

Tynkkynen, S., Santonen, T., Stockmann-Juvala, H. (2015). A comparison of REACH-derived no-effect levels for workers with EU indicative occupational exposure limit values and national limit values in Finland. Annals of Occupational Hygiene 59, 401-415.

US EPA, Environmental Protection Agency. (1997). Exposure Factors Handbook Vol. I. General Factors, U.S. Environmental Protection Agency Washington DC.

Waters, M., McKernan, L., Maier, A., Jayjock, M., Schaeffer, V., Brosseau, L. (2015). Exposure estimation and interpretation of occupational risk: enhanced information for the occupational risk manager. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S99-111.

Wheeler, M.W., Park, R.M., Bailer, A.J., Whittaker, C. (2015). Historical context and recent advances in exposure-response estimation for deriving occupational exposure limits. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S7-S17.

WHO, World Health Organization. (2009). Environmental Health Criteria 240, Principles and Methods for the Risk Assessment of Chemicals in Food, IPCS International Programme on Chemical Safety; World Health Organization, Geneva, Switzerland.

List of Tables

REPORT 2:
Benchmark Dose Modelling

RESEARCH PROJECT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Summary

Benchmark dose (BMD) modelling is the state of the science for the determination of Point of departures (PODs) for risk assessment. Technically it is done with software or online applications that fit flexible mathematical models (or a group of models) to experimental data. The BMD method is applied to the data of a specific endpoint. It determines a mathematical function that describes the dose-response relationship of the experimental data. The effect size considered is called benchmark response (BMR) or critical effect size (CES). The corresponding benchmark dose (BMD) describes the dose with a predefined additional effect compared to the control. Confidence intervals express the uncertainty due to sampling and/or measurement error. The benchmark dose lower bound (BMDL) describes the lower bound of the (in general, one-sided) 95th confidence limit of the BMD, the BMDU the upper limit. Both the BMD and the BMDL are used as POD for further assessments, but the BMDL is recommended as it considers the uncertainty pertaining to the BMD.

Three relevant guidance documents for the application of the BMD approach in risk assessment are available from different institutions: from the EFSA, the US EPA and from WHO IPCS. With the different guidance documents come different software tools or online-tools that all do the same job: fit mathematical models to the data and calculate BMDS, BMDLs and BMDUs.

In a first step the risk assessor has to decide which data should be selected for BMD modelling. A systematic search and screening step is necessary at first followed by the identification of reliable and relevant studies that can be subject to BMD modelling.

Next, the BMR for the effect considered has to be selected by the assessor and set in the tool used for the analysis. The selection of the BMR depends on the type of data selected. In case of quantal data (describing the frequency of occurrence of a feature in the examined group) a BMR of 10% (extra risk) is recommended by all three guidance documents.

For continuous data (describing a feature that is measured on a continuous scale, like body weight) the guidance documents do not share the same opinion. Recommendations go from 5% change in the mean response compared to the effect level in the control group (EFSA recommendation) to the use of 1 standard deviation (SD) as suggested by the US EPA. However, it can be concluded that the BMR for continuous data should be modified based on toxicological or statistical considerations.

During the next step models are selected (by the modelling tool or the assessor) that are fit to the data. The functional equations are determined by several parameters. Therefore, the model fitting process applied during dose-response modelling defines the parameters in such a way that the models come as close as possible to the observed dose-response data.

Over the last decade it has been shown that the best way to account for model uncertainty is model averaging. For model averaging the individual model results are combined based on their goodness of fit. This means that a better fitting model weighs more in model averaging than a model which fits less good to the data. In the end, the BMD and BMDL are calculated based on the “average model”.

If model averaging is not available or undesired for a special case, BMD and BMDLs can be selected as described in the EFSA guidance.

The BMD or BMDL obtained during this process can be used as POD for further risk assessment.
In the guidance documents for the derivation of OEL values the benchmark approach is described in most cases only as an alternative for the extrapolation from LOAEL to NAEL. Background information on the approach is often limited and guidance on the application is missing.
This contradicts the description of this method as “state of the science” for determining a point of departure (POD) for risk assessment.

Abbreviations

AEL	Acceptable Exposure Levels
AAEL	Acute Acceptable Exposure Levels
AF	Assessment factor
AGS	Ausschuss für Gefahrstoffe
AIC	Akaike information criterion
ANSES	Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail
AOEL	Acceptable Operator Exposure Levels
AAOEL	Acute Acceptable Operator Exposure Levels
APROBA	Approximate probabilistic analysis
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
BBMD	Bayesian Benchmark Dose
BMD	Benchmark dose
BMDL	Benchmark dose lower bound
BMDU	Benchmark dose upper bound
BMR	Benchmark response
BMDS	Benchmark dose software
BOELV	Binding occupational exposure level values
BPR	Biocidal products regulation
BS	Bootstrapping
CDS	Cumulative distribution function
CES	Critical effect size
CSAF	Chemical-specific adjustment factors

DFG	Deutsche Forschungsgesellschaft
DMEL	Derived minimal effect level
DNEL	Derived no effect level
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
ED10	Effective dose 10% (dose corresponding to a 10% increase in an adverse effect, relative to the control response)
EFSA	European Food Safety Authority
GM	Geometric mean
GSD	Geometric standard deviation
GV	Guidance value
IPCS	WHO's International Programme on Chemical Safety
IRIS	Integrated Risk Information System
LOAEC	Lowest observed adverse effect concentration
LOAEL	Lowest observed adverse effect level
MAK	Maximale Arbeitsplatzkonzentration
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MCRA	Monte Carlo Risk Assessment
MPPD	Multiple path particle dosimetry (model)
NAEC	No adverse effect concentration
NAEL	No adverse effect level
NOAEC	No observed adverse effect concentration
NOAEL	No observed adverse effect level

OEL	Occupational exposure limit
PBPK	Physiology-based pharmacokinetic (model)
PDF	Probability density function
POD	Point of departure
PPP	Plant protection products
PROAST	Dose-response modelling software by RIVM
QSAR	Quantitative structure activity relationship
RAC	Committee for Risk Assessment
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals,
RfD	Reference dose
RIVM	Dutch National Institute for Public Health and the Environment
SC	EFSA's Scientific Committee
SCOEL	Scientific Committee on Occupational Exposure Limits
STEL	Short-term exposure limit
SD	Standard deviation
TD	Toxicodynamics
TK	Toxicokinetics
TRGS	Technische Regeln für Gefahrstoffe
US EPA	Environmental Protection Agency in the US
WHO	World Health Organisation

1 Introduction

1.1 What is benchmark dose (BMD) modelling?

Benchmark dose modelling (or dose-response modelling) means fitting flexible mathematical models (or a group of models) to experimental data. This method is the state of science for determining a point of departure (POD) for risk assessment (Haber et al. 2018).

The BMD method is applied to the data of a specific endpoint. It determines a mathematical function that adequately reflects the dose-response data of the experiment (see **Figure 1-1**). The benchmark dose (BMD) describes a dose with a predefined additional effect compared to the control. The effect size related to the BMD is called benchmark response (BMR) or critical effect size (CES). Confidence intervals express the uncertainty due to sampling and/or measurement error. The benchmark dose lower bound (BMDL) describes the lower bound of the (in general, one-sided) 95th confidence limit of the BMD, the BMDU the upper limit. Both the BMD and the BMDL are used as POD for further assessments.

The following **Figure 1-1** schematically depicts the information described so far. It is a generalised illustration applicable for quantal and continuous data (see section 2.1 for details on quantal and continuous data).

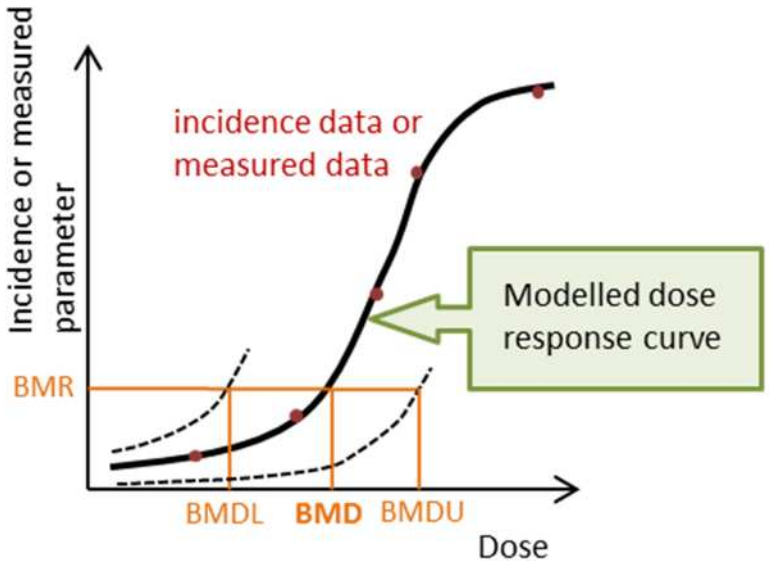


Figure 1-1 Schematic illustration of a BMD modelling. The important descriptors BMD, BMDL, BMDU and BMR are depicted.

1.2 Comparison of the BMD approach and the NOAEL-approach

Up to now the NOAEL (No observed adverse effect level) most often is used as POD for toxicological assessments. The NOAEL is the highest dose tested in experimental settings with no statistically significant increased occurrence of adverse effects.

In this section advantages and disadvantages of using the BMD approach are discussed (Muri et al. 2009; Schneider and Kaiser 2012; Slob 2014a, b; US EPA 2012).

Advantages:

- The NOAEL is found by comparing the observed effect in a dose group with the background effect (i.e., observed value in the control group). The highest dose tested with no statistically significant difference to the control is the NOAEL; the lowest dose tested with a significant difference is termed LOAEL (Lowest observed adverse effect level). Choosing the NOAEL as POD results in disregarding dose-response information above the NOAEL. In contrast, the BMD approach uses all dose-response information in the determination of the BMD/BMDL.

- In addition, the NOAEL is always one of the selected experimental doses, so it depends numerically on the choice of doses. The BMD approach is less dependent on dose selection and spacing of the experimental study.

- Datasets with few animals per dose group or with high variability will make it difficult to prove statistically significant differences. In consequence, doses with relevant effect levels might be identified as NOAELs. With the NOAEL approach thus poor study designs are “rewarded” with higher PODs. By using the BMDL, the BMD approach takes into account the uncertainty and variability of the data. Higher uncertainty results in lower BMDL values.

- If for a given dataset no NOAEL can be identified (all tested dose groups show significant effects), then extrapolation from a LOAEL to the NAEL (no adverse effect level) would be required, which introduces additional uncertainty into the assessment. If there are dose-response data informing the lower effect range, the BMD approach can be applied without further extrapolation step, thus avoiding the uncertainty of LOAEL-NAEL extrapolation.

- Kalantari et al. (2017) could show that study designs with more dose groups around the targeted BMD provided an BMD estimate that was slightly better than the one coming from a conventional study design. The authors conclude that in situations with a clear dose-response fewer animals receiving high doses could be achieved. This minimises overall animals’ distress.

- As Slob described in his twin papers in 2014 (Slob 2014a, b) the BMD approach allows to combine similar datasets for the same chemical in a single analysis (e.g.

both sexes), which can result in a reduction of animals while the precision is kept. In addition, it allows a quantification of the precision of the BMD estimate (which the NOAEL approach does not).

Disadvantages:

- The BMD approach is more complex and time-consuming in its application than the NOAEL approach. The user has to gain some background information on the method and get used to the selected software.

- It is necessary to define for example the set of models which should be applied to the data or the BMR for quantal and continuous data. This requires an agreement and solid guidance documents that are accepted in the scientific community and can serve an orientation for risk assessors.

- Current study designs (e.g. of repeated-dose studies according to OECD guidelines) are not optimised for use with the BMD approach. Typically, three dose groups plus a control group with a number of animals allowing to statistically detect differences are used. With this low number of doses, the use of the BMD approach is jeopardised in case you cannot use data from all dose groups. Examples for such a situation are:

    - the highest dose cannot be evaluated due to increased mortality;

    - the lowest two dose groups show no effect, only the top dose determines the course of the dose-response curve.

  In such cases the shape of the dose-response curve is insufficiently described, introducing uncertainty to the BMD modelling. For the application of dose-response modelling, use of more dose groups with fewer animals per dose would be advantageous (Slob 2014a).

2 Methodological principles

2.1 Types of dose-response data

For BMD modelling mainly two different types of data have to be distinguished.

- **Quantal data (dichotomous data)**

  This data type describes the frequency of occurrence of a feature in the examined group (also called incidence).

  Example:

Dose group	**Number of animals**	**Number of animals showing an effect (e.g., kidney failure)**
0 (control)	20	1
1	20	3
2	20	7
3	20	12

- **Continuous data**

  Continuous data describe a feature that is measured on a continuous scale, like body weight or concentration of protein in urine. Typically, continuous data are characterised by the group average value and its standard deviation.

  Example:

Dose group	**Body weight (mean ± SD)**
0 (control)	34 ± 4 g
1	33 ± 3 g
2	30 ± 2 g
3	25 ± 3 g

- **Categorical data (ordinal data)**

  In case of a third type of data, so-called categorical (or ordinal) data, test results are divided into groups, often according to a qualitative classification by severity. Categorical classifications are applied for example to express differences in the

results of histopathological examinations that cannot be described numerically (Chen and Chen 2014).
Example: Histological changes, divided into the groups "none, mild, medium, high".
Please note that quantal data are equivalent to ordinal data in case of only one severity group.

With the PROAST software, and the EFSA web tool BMD/BMDL values can be calculated for categorical data (Varewyck and Verbeke 2017). Input data has to be presented graded with severity scores (0, 1, 2,…) for increasing severity with “0” meaning “normal” (RIVM 2019).

According to Davis et al. (2011) the US Environmental Protection Agency (US EPA) planned to include the CatReg software to perform categorical regression analysis in the Benchmark Dose Software (BMDS). However, in the newest version of BMDS available (version 3.2) the categorical regression software CatReg is not yet included.

The following table described the differences between continuous, categorical, and quantal data. Continuous data can be transformed into categorical or quantal data if the effect can be categorised or a cut-point for affected/not affected can be set. Therefore, the amount of information provided by the data decreases from left (continuous data) to right (quantal data).

Continuous data **value (arbitrary unit) per individual (all in same dose group)**	**Categorical data:** **<=1 low** **>1 medium <=3** **>3 high**	**Quantal data:** **cut-point 3**
0.3	low	not affected
0.7	low	not affected
1.5	medium	not affected
2.3	medium	not affected
3.9	high	affected
4.2	high	affected
4.8	high	affected

- **Nested data**

A fourth type of data are nested data; they are usually obtained in developmental toxicity studies and refer to individual responses within an experimental unit (e.g.,

litter). Models for nested data are available in BMDS and the PROAST-based web tools.

The data types differ in the models used and in the selection of the BMR (see sections 2.5 and 2.6).

2.2 Available tools for BMD modelling

In general, regression analyses and modelling of dose-response data can be done with many different statistical programs (e.g., SAS®)[1]. These programs typically include several alternative methods, e.g., least-square and maximum-likelihood model fitting methods. However, non-statisticians are advised to use programs that have been developed specifically for the use in toxicology.

The following specific software tools are available. The tools are individually presented in the following sections:

- PROAST package for R software

  The R software and the PROAST package are freely available and can be installed on any computer.

- Benchmark Dose Software (BMDS), current version 3.2 (version 3.1 was used for example calculations in this project)

  In September 2018 BMDS 3.0 was released by the US EPA and several small bug-fixes were included in version 3.1 (released in February 2019). Version 3.2 was released in August 2020. BMDS is also free of any charges and can be installed in connection with Microsoft Excel®.

The following web tools are currently available:

- PROAST web tool (https://proastweb.rivm.nl/)

  No registration required, free to use for everybody

- EFSA web tool (https://r4eu.efsa.europa.eu/)

  Registration (E-mail/password) required, can be used freely after registration

- Bayesian Benchmark Dose (BBMD) Analysis System (https://benchmarkdose.com/)

  Registration (E-mail/password) required, can be used freely after registration

Differences in the software tools are presented in the individual sections in sections 2.5, 2.6, and 2.7 and in the following sections.

[1] http://www.sas.com/

2.2.1 PROAST software

PROAST is a software package that has been developed by the Dutch National Institute for Public Health and the Environment (RIVM) for the statistical analysis of dose-response data. PROAST has been originally developed as a package in R. Before PROAST can be run on a computer, the R software has to be downloaded and installed. A detailed description on how to set up the environment to run PROAST properly is given in the PROAST manual that is included in the download package (available here: https://www.rivm.nl/documenten/proast700)
The current version of the PROAST package is 65.5[2]. RIVM is constantly working on the optimisation of the package and new versions are available once in a while.
In the current EFSA guidance on applying the BMD approach (EFSA Scientific Committee et al. 2017) major differences can be found compared to the previous version (EFSA 2009a). These changes are reflected in the PROAST packages starting with version 62.0.
For the major differences to previous PROAST versions resulting from the new EFSA guidance (see section 2.3.1).

2.2.2 BMDS[3]

The BMDS was developed by US EPA in the nineties. Over the last 20 years several updates and new versions of the program have been published. The newest version of BMDS, version 3.2 was published August 2020 and is available at: https://www.epa.gov/bmds/benchmark-dose-software-bmds-version-32-download .
The program BMDS is consistent with the US EPA´s Risk Assessment Forum's Benchmark Dose Technical Guidance Document (US EPA 2012) and the technical guidance for choosing the appropriate stage of a Multistage model for cancer modelling[4]. In BMDS version 3.0 Bayesian versions of all models are included.
For the newest version of BMDS (US EPA 2018) Microsoft Excel® is required.

2.2.3 PROAST web tool

The PROAST web tool is available at the following website: https://proastweb.rivm.nl/. This online version may be very useful for users who do not want to install the R software and only do BMD modelling occasionally. The usual dose-response analysis of toxicity data can be done with this web tool. However, it has to be kept in mind, that the web application does not include all functionalities of the R version of PROAST. The current version of the PROAST web tool is based on PROAST version 66.39 (as at 01.07.2019).
Additional information and manuals are available on the PROAST homepage (see 2.2.1).

[2] Checked on 01.07.2019
[3] https://www.epa.gov/bmds/about-benchmark-dose-software-bmds
[4] https://cfpub.epa.gov/ncea/bmds/recordisplay.cfm?deid=308382

2.2.4 EFSA web tool

The EFSA web tool can be accessed here: https://r4eu.efsa.europa.eu/. After registration the user has access to an online application that implements statistical methods for BMD modelling using the R-package PROAST (version 66.38, status quo on 01.07.2019) and a user friendly surface (EFSA 2018). The applicability and functionalities of the R version of PROAST and the EFSA web tools are comparable.

2.2.5 BMD Analysis System

The Bayesian Benchmark Dose (BBMD) Analysis tool is available here: https://benchmarkdose.com/. It was developed by Shao and Shapiro (2018).
The BBMD Analysis tool has an implementation of Bayesian inference for benchmark dose estimation. The Bayesian framework provides the possibility to include prior information through prior distribution of model parameters. This is especially useful for poor-quality data to enhance their reliability. In addition, the BBMD Analysis System offers the possibility of a probabilistic risk assessment. For details see Part 3 (Probabilistic hazard assessment).

2.3 Available guidance documents

The two main documents available in the published literature describing the use of the BMD approach for toxicological risk assessment are:

- EFSA guidance on the use of the benchmark dose approach in risk assessment (EFSA Scientific Committee et al. 2017)

- US EPA guidance "*Benchmark dose technical Guidance*" (US EPA 2012)

For both documents previous versions are available. In the following two sections these guidance documents are presented.

Further, Chapter 5 of the Environmental Health Criteria document 240 was recently updated (WHO 2020). The final conclusions were discussed in a meeting in March 2019 in Geneva. In this Chapter 5 "Dose-response assessment and derivation of health-based guidance values" existing approaches to dose-response modelling were reviewed and generic approaches, with specific focus on harmonising approaches used in the USA and in Europe, independent of the tool used, were developed. It is briefly summarised below (section 2.3.3).

2.3.1 EFSA BMD guidance

In 2005 EFSA's Scientific Committee (SC) was requested by EFSA to assess existing information on the application of the BMD approach as an alternative to the NOAEL approach. Therefore, in 2009 SC published a guidance on how to use the BMD approach for analysing dose-response data from experimental and epidemiological studies (EFSA 2009a).

In 2015 the SC decided to update the guidance, especially the part on how to apply the BMD approach in practice.
In the current version of the EFSA guidance on applying the BMD approach (EFSA Scientific Committee et al. 2017) major differences can be found compared to the previous version (EFSA 2009a):

- The Akaike information criterion (AIC) was selected instead of the log-likelihood for characterisation of the goodness of fit for the different mathematical models applied.

  The Akaike information criterion (AIC) gives an estimate for the relative quality of a model. In general, a "good model" has minimum AIC among the other models.

- For continuous data models 3 and 5 from the Hill model families are considered. From these two models the one with the lower AIC is used for BMD calculation of BMDL.

- Model averaging is recommended as the preferred method for calculating the BMDL

With the help of a flow chart, the user is lead through a BMD analysis (**Figure 2-4** in section 2.7.1).

The SC points out that it is always recommended to report the BMD confidence interval with the BMDL (lower limit) and the BMDU (upper limit). The BMDL is used as POD and the BMDU/BMDL ratio reflects the uncertainty of the BMD estimate (EFSA Scientific Committee et al. 2017).

As already mentioned before, the current version of PROAST (65.5) and the PROAST-based web tools are based on the latest version of the EFSA guidance.

2.3.2 US EPA BMD guidance

The BMD guidance from the US EPA was prepared by a technical panel under the auspices of U.S. EPA's Risk Assessment Forum. During a peer review process, public consultation in the year 2000, comments and experiences from users and scientists the current version of the guidance was developed (US EPA 2012).
The guidance document is intended:

- to provide guidance for the members of the US EPA and the outside community on the application of the benchmark dose approach,

- to inform about preferred computational algorithms in the available software in order to allow users to make an informed choice in the selection of that software,

- to give information on the design of studies intended to be evaluated with BMD methods.

In addition to the guidance document the user manual for BMDS version 3 (US EPA 2018) can be used as a source of information since the guidance published in 2012 is not up-to-date on all areas.

2.3.3 Chapter 5 “Dose-response assessment and derivation of health-based guidance values” of EHC 240

Risk assessors, mathematicians, and statisticians from many countries and institutions worked together to develop common views and to promote use of dose-response modelling. Specifically, methodological differences between approaches developed in the USA (for use of the BMDS software by the US Environmental Protection Agency) and in the EU (by EFSA with PROAST-based tools) were discussed. The following conclusions were drawn (WHO 2020):

- There are continuous efforts to provide a similar set of models in both BMDS and PROAST-based tools, increasing the comparability of approaches and results; there is the common understanding that the type of models used should be of minor importance for determining the quantitative output.

- A priori parameter restriction should be minimised

- With the implementation of model averaging for both quantal and continuous data model selection will be of minor importance, as all models (but to differing degrees depending on their ability to describe the empirical data) are used (see section 2.6.5 for more details).

- For quantal data 10% incidence is a generally accepted level to identify the POD.

- For continuous data the benchmark response is more difficult to define. It was concluded that an expert decision should be taken to determine a BMR which characterise the borderline to adversity, “considering the type and severity of the effect, the background variability and the MOA leading to the effect” (WHO 2020). This approach is preferred over statistical approaches (e.g., BMR = background plus two standard deviations) or a fixed relative deviation from background (e.g., 5%).

2.4 Selection of critical studies and endpoints

If a toxicologist wants to perform a BMD analysis for derivation of a POD for risk assessment, available studies have to be systematically searched and screened in a first step. It seems needless to say that a complete review of the toxicity data is necessary in case the results of the BMD analysis are used for a toxicological risk assessment or the derivation of limit values of any kind. Identified key studies of high reliability and relevance will then be subjected to dose-response analysis.

As pointed out by EFSA Scientific Committee et al. (2017) an important step during hazard identification is consideration of dose dependency of the observed effects. In general, this is done by visual inspection and statistical evaluations of the data. In

case of a huge amount of relevant (statistically significant) data, visual inspection may be used to select those datasets with effects at lower doses. In an ideal approach selection of the critical datasets should be based on BMD analysis of all relevant datasets. Of course, the critical viewpoint of a toxicologist should always be applied during this procedure.

At the end of this procedure potentially critical effects are identified that should be analysed in more detail. In a recent review (Haber et al. 2018) present as a “rule of thumb” to model all endpoints with a LOAEL within a factor of 10 of the lowest LOAEL in the database.

2.5 Selection of BMR (benchmark response)

The BMR is a specific value of the effect size which is selected by the assessor and set in the tool used for the analysis. It is used for estimating the associated dose which is then named the corresponding BMD (benchmark dose). Small numbers or letters behind the word “BMD” or “BMDL” indicate which BMR was selected (e.g., BMD_{05} or $BMDL_{10}$ for a BMD at the 5% incidence level or a BMDL at the 10% incidence level, resp.). Details for quantal and continuous data are presented in the following two sections.

For quantal data the guidance documents of EFSA and US EPA are largely consistent. For continuous data the two institutions went quite contradicting ways. As Haber et al. (2018) pointed out in their review article comparing EFSA´s and US EPA´s recommendation: “*Clearly, the definition of the BMR includes judgement elements of science policy*”.

2.5.1 Quantal data

For quantal data, the BMR is defined as the increase in frequency of affected individuals compared to the control.

There are two ways expressing the exposure-related increase in frequency: "additional risk" and "extra risk". The difference between these two risk terms is based on the way how the incidence of the control group ("background incidence" is considered (Sand et al. 2008):

- The "additional risk" represents the absolute exposure-related increase of the incidence compared to the control (see equation 2.1).

$$\text{"additional risk"} = \mathrm{P}(\mathrm{d}) - \mathrm{P}(0) \qquad (2.1)$$

- For the "extra risk" this increase is relative to the proportion of the control group that did not show an effect (resulting from 100% minus background incidence, see equation 2.2).

$$\text{"extra risk"} = (P(d) - P\ (0))/(1 - P\ (0)\ ) \quad (2.2)$$

With
P (d) = effect level at dose d and
P (0) = effect level in control group

In case, the control group is without any effect (P (0) = 0) "extra risk" and "additional risk" will not differ. If the background incidence is > 0, the “extra risk” will always be higher than the "additional risk".

In the scientific literature the discussion about establishing a BMR for quantal data goes back to the beginnings of the BMD method (Allen et al. 1994a, b; Faustman et al. 1994; Kavlock et al. 1995; Kodell 2009). Allen et al. (1994a; b) examined the relationship of BMDL to NOAEL in 246 studies with 1825 endpoints on developmental toxicity. On average, the NOAELs were two times higher than the $BMDL_{10}$. Since no comparison was made with the BMD, a statement on the incidence at the NOAEL is difficult. However, taking the confidence intervals in developmental toxicity studies, an incidence level of 5 to 10% is expected (BMD/BMDL presumably > 2). Using models specifically designed to model developmental toxicity data, NOAEL values were similar to $BMDL_{05}$ (Allen et al. 1994b). The recommendation of a workshop on the application of the BMD method accordingly was to use the $BMDL_{05}$ or $BMDL_{10}$ (Barnes et al. 1995).

In a review article by Sand et al. (2008) other systemic endpoints than developmental toxicity were evaluated. The authors concluded that NOAEL values ranged between 5% and 10%. In an evaluation from 2011 (Sand et al. 2011) investigated chronic rat and mouse data generated by the National Toxicology Program (NTP). In 1183 datasets a median for the incidence of effects of 10% at the NOAEL was found. This corresponds to a BMD_{10} which was also calculated by the authors.

In an evaluation performed by Wignall et al. (2014) 800 dose-response datasets for 352 chemicals were evaluated with BMDS in a standardised way (quantal and continuous data). For both data the authors could show that the ratio between $BMD_{10\ or\ SD}$ and NOAEL values was < 2, and the ratio between $BMDL_{10\ or\ SD}$ and NOAEL was even lower than 1.

In their guidance document EFSA recommends to use a BMR of 10% incidence (extra risk) for quantal data (EFSA Scientific Committee et al. 2017), also referring to the above mentioned studies. EFSA does not distinguish between animal and epidemiological data in its recommendation. However, in a document from some years back EFSA (2011) states that "*No default BMR value is proposed for human data because of the difference in the types of studies and quality of data*."

The US EPA (2012) recommends using a BMR of 10% (extra risk) for quantal data to allow comparisons across chemicals and endpoints. The authors point out, that there may be cases where other response levels could be used, if this is supported by the statistical and biological characteristics of the data (e.g., 5% for frank effects or >

10% for early precursor effects). For epidemiological studies a BMD_{10} can often imply an upward extrapolation. In this case a BMD_{01} can be selected to avoid this situation (US EPA 2012).

2.5.2 Continuous data

For continuous data, the BMR can be defined in several ways.

Bokkers and Slob (2007) reanalysed a large number of NTP studies and showed that the $BMDL_{05}$ (5% change compared to the control) on average was close to the NOAEL values for the same endpoints.

According to Sand et al. (2006) a BMR of 5% to 10% (change compared to control) for continuous data may be appropriate from a risk assessment point of view.

Zeller et al. (2017) analysed historical vehicle control data of standard in vivo genotoxicity tests (micronucleus tests, comet assays, transgenic rodent or Pig-a assays) with statistical methods. The authors conclude that BMR values which were derived from the standard deviation (SD) of the respective study´s control group were highly variable for the same endpoints. They propose instead, to use a complete set of historical control data for quantification of the variability of vehicle controls. As a pragmatic approach Zeller et al. suggest to use the BMR_{1SDthe} (mean and SD of historical data after excluding the uppermost 5% data points).

EFSA recommends defining the BMR as a percentage of change in the mean response compared to the effect level in the control group ("background response"). The recommended default value is a BMR of 5%. This means that a deviation of 5% compared to the value of the measurement parameter in the control group is considered to be a critical effect (EFSA Scientific Committee et al. 2017). In the EFSA and the PROAST web tools the BMR is defined as explained here with a suggested value of 5%. Nevertheless, according to the EFSA guidance the BMR might be modified based on toxicological or statistical considerations.

This approach is transparent and understandable. However, it does not consider the variability of the data nor the nature and severity of the effect. For effects with high variability among animals, the differences between individual unexposed animals may exceed 5%. It can be questioned whether the critical effect size should be within the range of the natural variability of the effect parameter. However, with the “5%-approach” a comparison among studies and populations that differ in within-group variation is possible (Slob 2017).
Almost two decades ago RIVM tried to define BMR for individual effects (Dekkers et al. 2001). Since the nature of a measured effect can be very different and a 5% change of an early precursor effect with borderline adversity is judged differently than a serious effect this individual approach was a reasonable idea. However, this approach was not pursued any further in the benchmark community.

The US EPA recommends to use a biological basis for the BMR which is significant for the respective effect, meaning, the effect can be regarded as adverse at this

value (US EPA 2012). In the absence of a biological basis the US EPA favours using one standard deviation (SD) from the modelled mean of the control to determine the BMR. This procedure has the advantage that the variability in the effect parameter is included in the definition of the BMR.
However, in this case, the associated BMD depends on one particular study (measurement errors, dosing errors, heterogeneity in experimental conditions etc.). In addition, with the “1 SD-approach” a translation of the BMD to an equipotent dose in populations with larger intra-individual variations (e.g. humans) is not possible (EFSA Scientific Committee et al. 2017).

The BMDS offers the following options for BMR in continuous data:

- Standard deviation (proposed by the US EPA)
- Relative deviation from mean values of the control group (EFSA proposal: 5%)
- Absolute deviation
- Point (means the response associated with the BMR will be a numerical value, specified by the user, indicating the response, or change in response, of interest)
- Hybrid-extra risk (percentage change between the estimated deviation of maximum and minimum values)

As pointed out by Slob (2017) it is still discussed in the benchmark dose modelling community which BMR should be used for continuous data; and as can be seen from the explanations above, EFSA and the US EPA differ in their recommendations what BMR should be used.
In a recent publication Slob (2017) analysed data of 27 biological parameters for maximum response (M) and within group standard deviation (s). Slob (2017) presents the theory on “effect size”, meaning that the critical effect size should be scaled based on quantitative properties which are related to the specific endpoint. What should be taken into account for this approach are the maximum response (M) and the within group standard deviation (s). Slob could show that the two parameters “M” and “s” are positively correlated. He suggests to scale the BMR to the maximum response or, as a surrogate, to the within groups variance. Slob concludes: “*…the theory presented here makes clear that the rationale behind the* BMR_{SD} *is adequate, but only when the response data are log-transformed, and only when the value of the SD (on log-scale) represents the typical value for that endpoint in the long run (i.e. averaged over different studies). The preferred way of dealing with the benchmark response, however, is by using a value of CES* [critical effect size] *that is adjusted to the "expressiveness" of the particular endpoint by using information on M and s, where s is the typical value over a range of different studies. In this way, the benchmark response covers the rationale behind the* BMR_{SD} *but with the crucial advantage that the scaled CES is expressed as a percent (or fold) change, which is biologically/toxicologically interpretable, while making the associated BMD suitable for extrapolation to an equipotent dose in the (median) human being (WHO 2014).*”

As mentioned above, in the updated Chapter 5 of the Environmental Health Criteria document 240 on dose-response modelling it was concluded that the BMR for continuous data should primarily be set using toxicological criteria for adversity of effects. The BMD should indicate the borderline where adversity of effects begins, considering the objective and definition of the guidance value to be derived.

2.5.2.1 Hybrid approach

The US EPA also mentions the so called “hybrid approach”. With this approach continuous data are expressed like quantal data by using the distribution of continuous data and an estimation of the incidence of individuals falling above or below a level considered to be adverse (Crump 2002; Crump 1995; Gaylor and Slikker 1990). For this approach a cut-off criterion must be defined, above which individuals are considered affected. Assuming a log-normal distribution for each dose, it is possible to calculate the prevalence of affected individuals at each dose (EFSA, 2009c). In **Figure 2-1** the hybrid approach is shown schematically.

The US EPA guidance (US EPA 2012) states on the hybrid approach: *“The result is an expression of the data in the same terms as that derived from analyses of quantal data. That is, the approach implicitly dichotomizes the data, retaining the full power of modeling the continuous data while obtaining results that permit direct comparison of BMDs and BMDLs derived from continuous and quantal data.”* The hybrid approach is not associated with data loss that is generally associated with the conversion of continuous to quantal data (Crump 2002).

The hybrid method is implemented in BMDS 3.1, but not in PROAST.

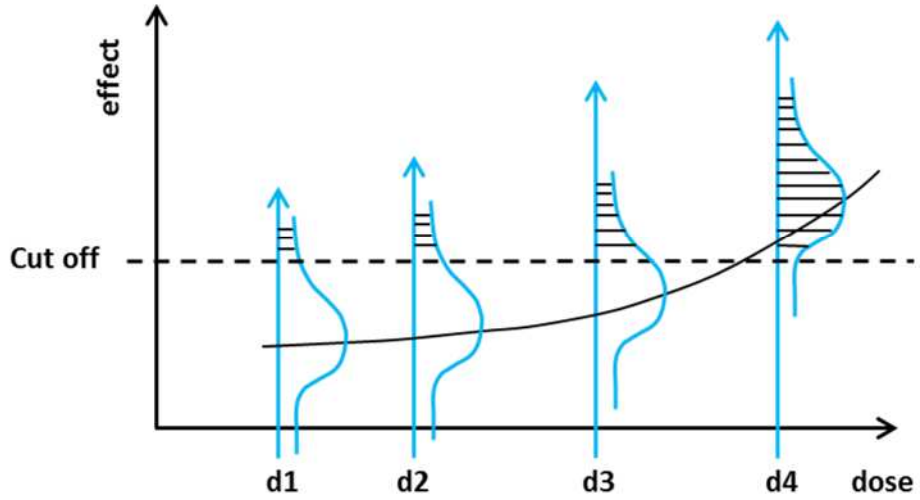


Figure 2-1 Schematic picture of the hybrid approach. For details see text. Figure modified from (EFSA 2009b).

2.5.3 Categorical data

According to the PROAST manual (RIVM 2019) the BMR (in this situation called CES) for categorical data is defined as followes:

"In PROAST [and presumably in the PROAST-based web tools] *the BMD for ordinal data is defined as the ED50 associated with a particular severity score, i.e. the severity score is considered as the BMR. In this way, each BMD may be regarded as the dose where the average animal would respond exactly on the borderline between the two consecutive categories. For instance, the ED50 associated with category 2 is defined as the dose where the average animal responds on the borderline between categories 1 and 2. This dose is analogous to the CED in continuous data, where it represents the dose at which the average animal's response is equal to the CES. In ordinal data the CES is defined as the transient from one severity category to the next. In the PROAST output CES = 0 means that the critical effect size is defined as transients between severity categories, and that the associated CEDs are ED50s."*

In the EFSA web tool a default of 5% is given as CES for categorical data.

2.6 Selection of models

Using a functional equation, the relationship between dose and response can be described. The "model" is a mathematical description of this relationship. Model selection should be data-driven and, ideally, models adequately describing the data should lead to very similar outputs as the experimental data fix the models within a narrow range.

Nevertheless, different mathematical models can lead to different BMDs and BMDLs. In order to minimise this uncertainty, a range of models should be applied to the available data and suitable models selected according to transparent criteria. The functional equations are determined by several parameters. Therefore, the model fitting process applied during dose-response modelling defines the parameters in such a way that the models come as close as possible to the observed dose-response data.
Recommended models for continuous and quantal data are presented in the following sections. However, as pointed out in the EFSA guidance, there are two special theoretical models that relate to both types of data (quantal and continuous): the "full model" and the "null model".

- The full model is a theoretical model that perfectly covers all data points; it may be used for evaluating the goodness of fit of any dose-response model.

- The null model is a theoretical model corresponding to a horizontal line and may be used for statistically evaluating the presence of a dose-response.

EFSA recommends fitting all available models (including the full and the null model) to the data. In the flow chart shown in **Figure 2-4** below this step is described with number **1**.

2.6.1 Quantal data

2.6.1.1 Available models

For quantal data, seven different models are recommended by EFSA (2017): Logistic, Probit, Log-logistic, Log-probit, Weibull, Gamma, Linearized (two stage) model and latent variable models (LVMs) based on the continuous models (see **Table 2-1**). According to the PROAST manual (RIVM 2019) LVMs "…*assume that the observed incidences originate from an underlying continuous response, which is not directly observed. Instead, each animal (experimental unit) is observed to have a response below or above a certain cut-off value, resulting in yes/no responses for all animals. This cut-off value is (normally) unknown, as it relates to the invisible latent variable. In fitting the model to the data, it can however be estimated […]. Thus, the models, defined for continuous response data, can be similarly used for quantal data. The LVMs similarly apply to ordinal data…"*
LVM were implemented in PROAST and PROAST-based web tools because they have been found to adequately describe quantal data in general. Especially, when model averaging (see section 2.6.5) is applied they should be included in the BMD analysis (EFSA Scientific Committee et al. 2017).

BMDS 3.1 offers the quantal models available in previous versions of BMDS (Logistic, Probit, Log-logistic, Log-probit, Weibull, Gamma, Hill, Multistage and Quantal linear, see **Table 2-1**). In addition, in version 3.1 of BMDS a Bayesian version of each model was added. These Bayesian dichotomous models are identical to the parametric models listed above. The main difference of these Bayesian models is the prior incorporation and usage of information into the analysis (US EPA 2018). This means that statistical methods are applied that assign probabilities or distributions to parameters based on prior data collection. According to Shao and Shapiro (2018) "*the Bayesian framework provides a way to incorporate prior information through the prior distribution of model parameters, which has great potential to enhance the reliability of dose–response modeling for poor-quality data, which may be the only data available for risk assessors in some situations. In addition, incorporating prior information may allow a reduction of the number of animals required for testing in future studies [..]. Second, owing to the distributional/probabilistic nature of this approach, a Bayesian dose–response modeling tool can facilitate probabilistic risk assessment, which is advocated by the scientific community".*

Table 2-1 Available quantal models in BMD-tools

Model name	**Model available in PROAST/PROAST based web tools**	**Model available in BMDS version 3.1**
Dichotomous Hill		X
Gamma	X	X
Logistic	X	X
Log-Logistic	X	X
Probit	X	X
Log-Probit	X	X
Multistage	X two stage model only	X BMDS may “auto select” the degree of the multistage model
Quantal-Linear		X
Weibull	X	X
Latent variable models (LVM)	X*	

* PROAST always applies model 3 and 5 as the LVM for the exponential and Hill model (see section on continuous data for more information).

For the quantal models available in PROAST and PROAST-based web-tools the following **Table 2-2** lists the mathematical equations.

Table 2-2 Model equations for the recommended quantal models in PROAST and PROAST-based-web tools according to EFSA Scientific Committee et al. (2017)

Model	**Model expression mean response (y) as function of dose (x)**
Logistic	$y = 1/(1 + \exp(-a-bx))$
Probit	$y = \text{CumNorm}(a + bx)$
Log-logistic	$y = a + (1-a)/(1 + \exp(-\log(x/b)/c))$
Log-probit	$y = a + (1-a)\ \text{CumNorm}(\log(x/b)/c)$
Weibull	$y = a + (1-a)\ \exp((x/b)^c)$
Gamma	$y = a + (1-a)\ \text{CumGam}(bx^c)$
LMS (two-stage) model	$y = a + (1-a)(1-\exp(-bx-cx^2))$
Latent variable models (LVMs)	These models assume an underlying continuous response, which is dichotomised into yes/no response based on a (latent) cut-off value that is estimated from the data

a, b, c, d: unknown parameters that are estimated by fitting the model to the data (see section 2.6.1.2 for details)
CumNorm: cumulative (standard) normal distribution function
CumGam: cumulative Gamma distribution function

2.6.1.2 Parameter restriction

In order to force a model not to have undesirable properties, model parameters can be restricted.

To understand better the restriction of parameters, the following **Figure 2-2** shows the three model parameters a, b and c for quantal data. The parameters are unknown and estimated by fitting the models to the data.

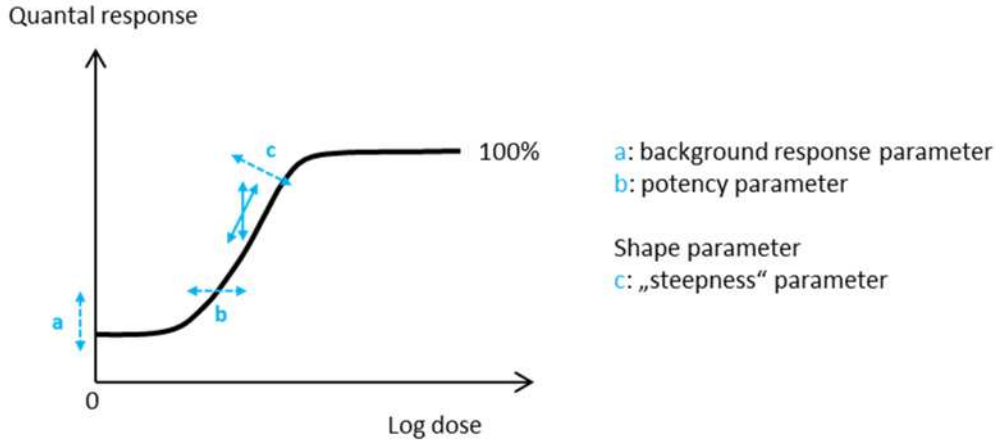


Figure 2-2 Adapted from EFSA guidance (EFSA Scientific Committee et al. 2017, page 22 Figure 6), showing the three model parameters a, b and c for quantal data. The dashed arrows show how the curve would change in case the respective parameter is altered.

In quantal data for example the “background” parameter “a” is restricted to be between 0 and 1, i.e. between 0% and 100%. In the following **Table 2-3** the differences in parameter restriction according to EFSA Scientific Committee et al. (2017) and US EPA (2012) are summarised. Please note that the US EPA gives different names to the individual parameters. For example, the background parameter is called “γ” in the US EPA guidance. For reasons of comparability the designations as given by EFSA were applied for all parameters.

In PROAST and PROAST-based web tools the “shape” parameter “c” reflecting the steepness of the curve is generally not restricted to be > 1 (it only has to be a positive value) as this may lead to artificially high BMDLs (EFSA Scientific Committee et al. 2017).

In BMDS most models can be applied restricted or unrestricted. As a default in BMDS, the Gamma, Log-Logistic, Multistage-and Weibull-models are restricted, whereas the Log-Probit- and Dichotomous Hill models are unrestricted. The Logistic, Probit and Quantal Linear models are only available in an unrestricted version (US EPA 2018). The US EPA points out that unrestricted models should only be applied if an acceptable fit is not obtained with any of the restricted models.

For the unrestricted Gamma model BMDS 3.1 sets the power parameter “α” (corresponds to parameter b in **Figure 2-2**) to be > 0.2 (US EPA 2018). The authors explain: “*If α < 1, then the slope of the dose-response curve becomes infinite at the control dose. This is biologically unrealistic, and can lead to numerical problems when computing confidence limits, so several authors have recommended restricting α ≥ 1. Note for the unrestricted Gamma model the α > 0.2 for numerical reasons.”* Apart from the Gamma model the other models are comparable to previous BMDS versions.

Table 2-3 Differences in parameter restriction for quantal data according to EFSA Scientific Committee et al. (2017) and US EPA (2012).

Model name	**Parameter restriction according to EFSA**	**Parameter restriction according to US EPA**
Dichotomous Hill	-	no information found
Gamma	$0 \leq a \leq 1$, $b > 0$, $c > 0$	$0 < a < 1$, $1 \leq b < 18$ (if restricted), $c \geq 1$ (if restricted)
Logistic	$b > 0$	$c > 1$ (if restricted)
Log-Logistic	$0 \leq a \leq 1$, $b > 0$, $c > 0$	$0 < a < 1$, $1 \leq b < 18$ (if restricted), $c > 1$ (if restricted)
Probit	$b > 0$	$1 \leq c \leq 18$ (if restricted)
Log-Probit	$0 \leq a \leq 1$, $b > 0$, $c > 0$	$0 < a < 1$, $1 \leq c \leq 18$ (if restricted)
Multistage	$a > 0$, $b > 0$, $c > 0$	no information found
Quantal-Linear	-	no information found
Weibull	$0 \leq a \leq 1$, $b > 0$, $c > 0$	$0 < a < 1$, $1 \leq b < 18$ (if restricted), $c \geq 1$ (if restricted)
Latent variable models (LVM)	$a > 0$, $d > 1$*	-

* see section 2.6.2.2 on details of parameters for continuous models.

2.6.2 Continuous data

2.6.2.1 Available models

For continuous data EFSA recommends two models from the Exponential family and two from the Hill family ("nested models") (EFSA Scientific Committee et al. 2017). See **Table 2-4** for details. One model with three and one model with four parameters respectively. In the previous EFSA guidance (EFSA 2009a) two other models were included from each family, but are no longer recommended by the Scientific Committee.

Two models are considered nested if one can emerge from the other by adding one or more parameters. Therefore the models from the Hill- and Exponential-family can have a different number of parameters.

In the version of BMDS used for this report (version 3.1), only the exponential model is available as a nested model for continuous data. The Hill model is available as a 4-parameter model. The other models specified in BMDS (Linear, Polynomial and Power) are not recommended by EFSA because of their possible unrealistic model properties (not monotonic, negative values).

Bayesian models are currently not available in BMDS for continuous models, however this is a feature planned for the future.

Table 2-4 Available continuous models in BMD-tools

Model name	**Model available in PROAST/PROAST based web tools**	**Model available in BMDS version 3.1**
Exponential family		
2-parameter model	-	X
3-parameter model	X	X
4-parameter model	X	X
5-parameter model	-	X
Hill family		
3-parameter model	X	-
4-parameter model	X	X
Polynomial	-	X
Power	-	X

For the continuous models available in PROAST and PROAST-based web-tools the following **Table 2-5** lists the mathematical equations.

Table 2-5 Model equations for the recommended continuous models in PROAST and PROAST-based-web tools according to EFSA Scientific Committee et al. (2017)

Model	**Model expression mean response (y) as function of dose (x)**
Exponential family	
3-parameter model	$y = a \exp(bx^d)$
4-parameter model	$y = a [c-(c-1)\exp(-bx^d)]$
Hill family	
3-parameter model	$y = a [1-x^d/(b^d + x^d)]$
4-parameter model	$y = a [1 + (c-1)x^d/(b^d + x^d)]$

a, b, c, d: unknown parameters that are estimated by fitting the model to the data (see section 2.6.2.2 for details)

2.6.2.2 Parameter restriction

The following **Figure 2-3** shows the four model parameters a, b c and d for continuous data.

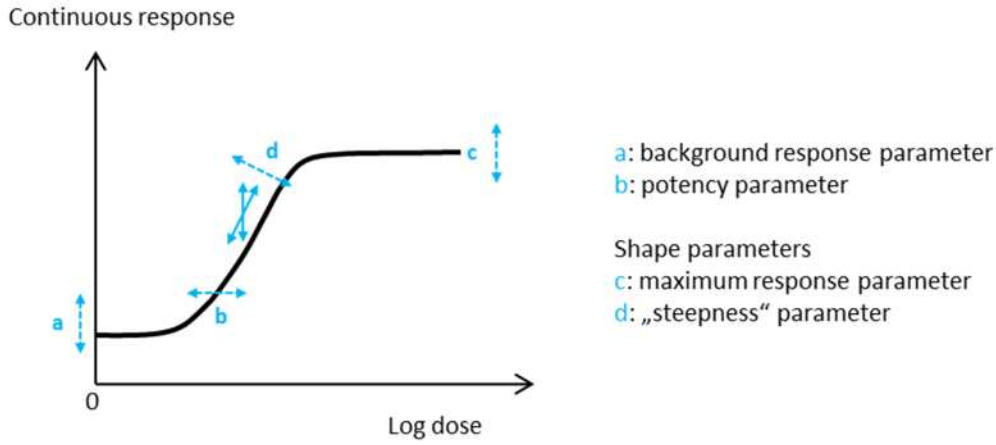


Figure 2-3 Adapted from EFSA guidance (EFSA Scientific Committee et al. 2017, page 22 Figure 6), showing the four model parameters a, b, c and d for continuous data. The dashed arrows show how the curve would change in case the respective parameter is altered.

In the following **Table 2-6** the differences in parameter restriction according to EFSA Scientific Committee et al. (2017) and US EPA (2012) are summarised. Please note again that the US EPA gives different names to the individual parameters. For reasons of comparability the designations as given by EFSA were applied for all parameters.

In BMDS the Hill, Polynomial and Power models can be run restricted or unrestricted The Exponential model can only be run restricted and the Linear model unrestricted. As explained already in the previous section on quantal data, the US EPA recommends using the unrestricted models only in case the restricted models did not result in an acceptable fit.

Table 2-6 Differences in parameter restriction for continuous data according to EFSA Scientific Committee et al. (2017) and US EPA (2012).

Model name	**Parameter restriction according to EFSA**	**Parameter restriction according to US EPA**
Exponential family		
2-parameter model	-	a > 0, d > 0
3-parameter model	a > 0, d > 1	a > 0, d > 0
4-parameter model	a > 0, b > 0, c > 0, d > 0	a > 0, d > 0 c > 1 (increasing data) 0 < c < 1(decreasing data)
5-parameter model	-	a > 0, d > 0, b > 1 c > 1 (increasing data) 0 < c < 1(decreasing data)
Hill family		
3-parameter model	a > 0, d > 1	-
4-parameter model	a > 0, b > 0, c > 0, d > 0	18 ≥ b > 1 (if restricted)
Polynomial	-	
Power	-	18 ≥ b ≥ 1 (if restricted)

2.6.3 Categorical data

In PROAST and the PROAST-based web tools categorical data are modelled with the latent variable models (LVMs). These kinds of models assume that increasing doses result in a gradual increase of the severity of the effect like in continuous data. For categorical data PROAST fits the exponential and the Hill latent variable models to the data.

2.6.4 Fitting the models

In an iterative process, the software tools modify the model parameters of each model until an optimal fit of the curve to the data is achieved. For fitting models, the software tools use statistical methods that adjust model parameters in such a way that the distance between the model curve and experimental data is minimized. This is done in a numerical way by maximization of the log-likelihood value. If the algorithm is able to find the maximum log-likelihood possible, the fit cannot be improved, and the software will report that the algorithm has converged.

2.6.4.1 Procedure according to the EFSA guidance

BMDS as well as PROAST and the PROAST based web tools apply the methods described in section 2.6.4 above. In the flow chart shown in **Figure 2-4** below this step is described with number **2**.

In order to compare the fit for different models, the AIC criterion is used. The AIC integrates the log-likelihood and the number of model parameters in one single value. The full model shows the smallest AIC and the null model the largest. The AIC can be used to check if there is statistical evidence for a dose-response relationship. The Scientific Committee from EFSA (2017) suggests that the AIC of any model should be lower than the AIC of the null model minus 2. In the flow chart shown in **Figure 2-4** below this step is described with number **3**.

In case that nested models were used, only one model is selected for each family of nested models. The model selected is the one that has the lowest AIC compared to the other models in the family. In the flow chart shown in **Figure 2-4** below this step is described with number **3a**.

In addition, the AIC of a fitted model should be no more than plus 2 larger than the full model's AIC. Based on these criteria models are selected or excluded for further analysis. In the flow chart shown in **Figure 2-4** below this step is described with number **3b**.

2.6.4.2 Procedure according to the US EPA guidance

The US EPA (2012) recommends selecting the models which should be applied to the data set based on the following criteria:

- Goodness-of-fit: p value of selected models should be > 0.1
- Reject models that do not adequately describe the low-dose data-points (visual inspection and examination of residuals (residuals that exceed 2 in absolute value warrant further examination of the model fit))
- If BMDL values of remaining models are sufficiently close, the model with the lowest AIC may be selected and the BMDL from this model is used as POD. If

two models share the lowest AIC, BMDLs from these models may be added up and divided by two.

- If BMDL values from the remaining models are not sufficiently close, judgement from a statistician is recommended. The lowest BMDL might be selected as POD.

- In case no useful results were obtained, the NOAEL/LOAEL approach should be considered instead.

Note:
The US EPA does not explain in the guidance document what “sufficiently close” BMDL estimates means. According to Haber et al. (2018) a factor of 3 was specified in the draft guidance.

The guidance document from the US EPA is from 2012. In the BMDS manual which was released in 2018 together with BMDS version 3.0 (US EPA 2018) a flow chart is included describing the criteria of BMDS for model recommendation. This process is done by the software. The following conditions are set by the software:

- If any of the following conditions are met, the model is moved to the “unusable bin”

  All data types
  1) Invalid BMD
  2) Invalid BMDL
  3) Invalid AIC

- If any of the following conditions are met, the model is moved to the “questionable bin”

  All data types
  1) BMD/BMDL ratio > 20
  2) Scaled residual of interest >2
  3) BMD 10x lower than lowest non-zero dose
  4) BMDL10x lower than lowest non-zero dose
  5) Degrees of freedom = 0, saturated model

  Continuous datasets only
  1) P-value < 0.05 for constant variance
  2) P-value < 0.05 for non-constant variance

  Continuous/quantal datasets
  1) Goodness of fit p-test < 0.05 (multistage cancer)
  2) Goodness of fit p-test < 0.1 (all other models)

- If any of the following conditions are met, the model is moved to the “viable” but with a warning

All data types
1) BMD/BMDL ratio > 5
2) BMDS output file included warning
3) BMD or BMDL higher than highest dose
4) BMD or BMDL 3x lower than lowest non-zero dose
5) BMDU not estimated

Continuous datasets only
1) Modelled response standard deviation > 1.5 x actual response standard deviation at control

- If none of the above-mentioned criteria applies the model is “viable”

2.6.5 Model averaging

Over the last decade it has been shown that the best way to account for model uncertainty is model averaging (Fang et al. 2015; Wheeler and Bailer 2009; Wheeler and Bailer 2007, 2008). For model averaging the individual model results are combined based on their goodness of fit. This means that a better fitting model weighs more in model averaging than a model which fits less good to the data. In the end the BMD and BMDL are calculated based on the “average model”.

Before model averaging was available, the BMD or BMDL used as POD for risk assessment was in general based on one single model. Today all available tools are able to perform model averaging for quantal data and it is highly recommended to apply this feature if quantal data are modelled. In one of the last updates of the EFSA-tool model averaging was included for continuous data as well (EFSA 2018; Slob 2018).
How to proceed in case that model averaging is not available, not working or due to any reason not desired is explained in section 2.7.

The EFSA guidance states: “*As the purpose of a BMD analysis is not to find the best estimate of the (true) BMD but rather to find all plausible values of the (true) BMD, given the data available, not only the best-fitting model but also the models resulting in a slightly poorer fit need to be taken into account. After all, it could well be that the second (or third, . . .) best-fitting model is closer to the true dose–response than the best-fitting model. This type of uncertainty is called ‘model uncertainty’, and implies that the BMD confidence interval needs to be based on the results from various models, instead of just a single (‘best’) model.”* (EFSA Scientific Committee et al. 2017).

In PROAST and PROAST-based web tools model averaging is based on the principles of Wheeler and Bailer (2007). The weight of the models is based on their AIC values, with better models getting larger weight. The POD is the model averaged BMDL which is estimated using parametric bootstrap methods.
In the flow chart shown in **Figure 2-4** below the step of model averaging is described with number **4**.

In BMDS version 3.0 Bayesian model averaging is currently only available for quantal data. The methods applied for this model averaging procedure are explained in detail in the manual document for BMDS version 3.0 (US EPA 2018). Background information on the Bayesian model averaging approach were published for example by Shao and Shapiro (2018), Shao and Gift (2014), Shao (2012), Kwon et al. (2016) and others.

2.7 Establishing the BMD confidence interval and setting the POD

A confidence interval expresses the uncertainty in a parameter estimate that results from sampling or measurement errors. The EFSA and US EPA guidance document both propose to use a (one sided) 95% confidence interval (the limit of a one-sided 95% confidence interval is equivalent to a two-sided limit of a 90% confidence interval). This means that the considered parameter has a 5% probability for being below the lower and a 5% probability to be above the upper confidence limit.

2.7.1 Procedure according to the EFSA guidance

According to the EFSA guidance (2017) the procedure as laid out **Figure 2-4** should be followed for establishing the confidence interval and the BMDL. Model averaging should be performed and the resulting BMDL can be used as POD (see section 2.6.5 above).
In case that model averaging is not available, not working or due to any reason not desired the BMDL has to be selected manually. In the flow chart shown in **Figure 2-4** below the step of selecting the BMDL/BMDU without applying model averaging is described with number **5**. Since modelling is limited to the experimentally observable area of the dose-response relationship, it is generally expected that the BMDLs obtained from the accepted models will depend little on model choice.

EFSA Scientific Committee et al. (2017) recommends the procedure laid out in the following flowchart (**Figure 2-4**):

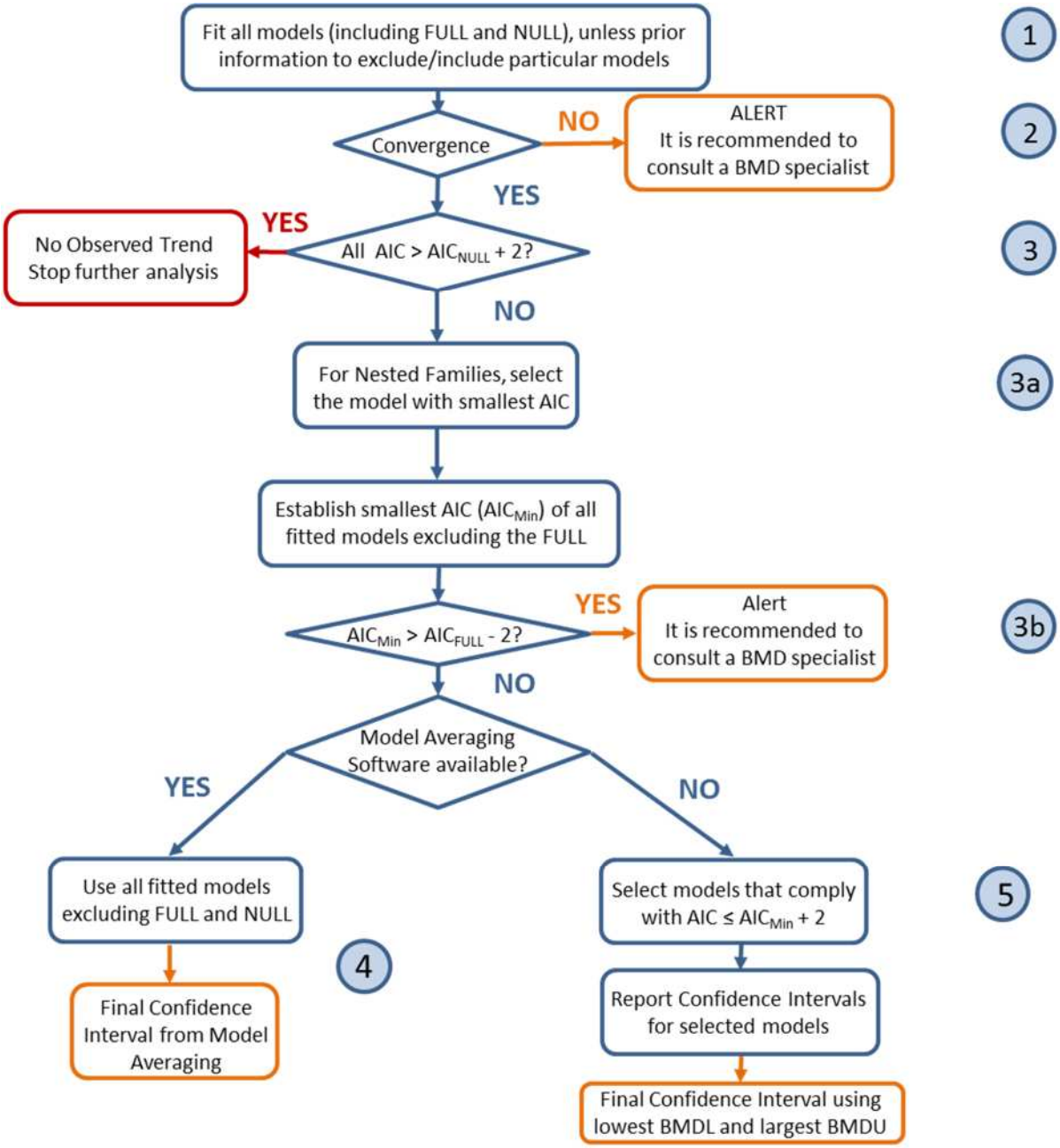


Figure 2-4 Flow chart adapted from EFSA guidance (EFSA Scientific Committee et al. 2017, page 28 Figure 8) on how to derive a confidence interval. The numbers in the blue circles refer to explanations in sections 2.6 and 2.7.

1) Perform steps 1, 2, 3, 3a and 3b according to flowchart

2) Apply model averaging (step 4 according to flow chart, as described in detail in sections 2.6.5)

3) In case no model averaging is performed continue with step **5:**

Based on the AIC models are selected or rejected. Models with an AIC lower or equal to the model with the minimal AIC + 2 are selected.
The confidence intervals for all models are reported and the lowest BMDL and the highest BMDU from all selected models are selected. This BMDL may be used as a POD for further assessments.

If accepted models result in very different confidence limits, this indicates that the dose-response data are not good enough and incapable of determining the models. According to EFSA (2017) it can be concluded: "*When the width of the combined BMD confidence interval is found to cover orders of magnitude, the BMDL could be orders of magnitude lower than the true BMD, had better data been available. Therefore, the resulting RP [reference point], and the HBGV [health-based guidance value] or MOE [margin or exposure] eventually derived from it, might have been much higher or larger, respectively.*"
In this case either other (better) data should be considered, or the available data should be reanalysed considering prior information on typical values of the shape parameters (for example from historical data): This could be done by restricting shape parameters or by applying prior distributions in a Bayesian approach.

Finally, if several endpoints were considered and BMD calculations performed, the lowest BMDL value is selected and defined as the BMDL value for the respective study. The lowest BMDL value selected from all endpoints and studies is usually the reference point for further calculations/extrapolations. In some cases this procedure is not optimal and a more elaborate approach may be necessary considering the biological meaning of the relevant endpoints and the consequences for the derived limit value (for details see EFSA guidance).

2.7.2 Procedure according to the US EPA guidance

According to the US EPA guidance (2012) the procedure as described in section 2.6.4.2 is followed automatically by BMDS. In case of quantal data model averaging can be performed and the resulting BMDL can be used as POD (see section 2.6.5 above).

If the range of BMDLs from models which are considered "viable" is < 3 it is recommended to use the BMDL from the model with the lowest AIC. Otherwise, it is recommended to use the lowest BMDL from all viable models. In contrast to PROAST and PROAST based web tools BMDS makes a recommendation for the model that should be taken as basis for the BMD/BMDL selection.

3 Discussion and conclusions

3.1 Area of application of the BMD method

3.1.1 General applicability for risk assessments

In the EFSA guidance (2017) the benchmark approach is explicitly mentioned for the derivation of health-based guidance values (HBGV) like acceptable daily intakes (ADIs), tolerable daily intakes (TDIs) or tolerable weekly intakes (TWIs). In the introduction of the EFSA guidance pesticides, additives and contaminates in food are listed for which the benchmark approach was already applied. The EFSA guidance points out that the BMD approach can be applied for non-genotoxic as well as for genotoxic substances.
It seems reasonable that EFSA is concerned with food contaminants or additives and applies the BMD approach for the derivation of limit values or Margin of exposure (MOE) calculations with regard to food safety.

In the US EPA guidance (2012) it is stated: "*These benchmark doses can then serve as possible points of departure (PODs) for linear or nonlinear extrapolation of health effects data and/or as bases for comparison of dose-response results across studies/chemicals/endpoints."* The US EPA does not focus on any specific area of regulation. The US EPA document is intended to provide guidance for a consistent use of the BMD method for a variety of purposes "…*including the determination of PODs for different types of health effects data, whether a linear or nonlinear low-dose extrapolation is used".* In general, the BMD method can be used on data for any kind of toxicological endpoint (including carcinogenicity).

According to a recently published review article by Haber et al. (2018) BMD modelling is currently considered the preferred approach for deriving PODs for risk assessments.

During a workshop in Brussels in March 2017 EFSA presented the updated guidance on using the BMD approach in risk assessment and the EFSA web tool for benchmark analysis[5]. Sixty experts from safety authorities across Europe, the US, Australia and New Zealand, Japan and the WHO participated in the meeting.
A broad consensus was reached concerning the following topics:

- The BMD approach is superior to the NOAEL approach
- Model averaging should be applied instead of single model analysis
- A variety of tools is available (including the web-based tools by EFSA and RIVM)
- All toxicological data should be considered

[5] https://www.efsa.europa.eu/en/events/event/170301-0 (assessed 23.1.2019)

- The documentation of BMD modelling should be transparent
- Training of assessors is needed
- An international platform should be established to share experience and views

Over the last 20 – 25 years the benchmark approach in toxicological risk assessment has been established, revised, and refined by numerous scientists. Today the BMD approach is the means of choice for deriving PODs in toxicological risk assessment. With the implementation of model averaging the reliability of the approach has increased tremendously. And the capabilities of the method have not been exhausted: different approaches for models averaging are discussed; model averaging for continuous data has to be implemented in some of the software tools, and discussions about relevant models or selection of BMR are ongoing.

However, the continued and extensive use of the NOAEL/LOAEL approach shows that many risk assessors are still not familiar with the method and/or consider the workload too high to apply it on a regular basis.

3.1.2 Applicability of the BMD approach for derivation of OEL values

In the ECHA guidance on information requirements and chemical safety assessment, Chapter R8 (Characterisation of dose-response for human health (ECHA 2012)) the benchmark dose concept is explicitly mentioned as an alternative to the NOAEL approach. ECHA highlights the importance of the BMD approach, if a NOAEL is not available and a LOAEL-NAEL extrapolation would have to be performed. Further guidance for example on the selection of the BMR or the models is not given in the R8 guidance document.

In the SCOEL methodology for the derivation of occupational exposure limits the benchmark dose is presented as an alternative to the NOAEL value (SCOEL 2013). However, no more information on the application of the BMD approach is given.

Since SCOEL was not reappointed in 2018 and ECHA´s Risk Assessment Committee (RAC) will from now on provide recommendations for occupational exposure limits to the European Commission[6] a methodology for the derivation of OEL values by RAC was released: In November 2018 a draft appendix to the ECHA guidance R8 was published (ECHA 2018) and was sent as a draft to CARACAL (ECHA 2019). This appendix provides guidance for deriving Occupational Exposure Limits. Under the section describing the assessment of carcinogenic substances, the BMD approach is listed as a method for deriving a POD. The authors of the appendix recommend the use of a BMD_{10} or the T25 method. For non-carcinogens, the BMD approach is not mentioned.

[6] https://echa.europa.eu/-/echa-to-provide-recommendations-for-occupational-exposure-limits (assessed 25.1.2019)

In the current version of the MAK- (Maximale Arbeitsplatzkonzentration) and BAT- (Biologischer Arbeitsplatz Toleranz-Wert) value lists from 2019, DFG also provides a methodology for the derivation of MAK-values (DFG 2019). In this methodology it is said that in the absence of a NOAEL no scientifically justifiable MAK-value can be proposed. The benchmark approach is not mentioned at this point. In the chapter describing the approach for (sensory) irritating substances the $BMDL_{05}$ or the $BMDL_{SD}$ are mentioned as alternatives in the absence of a NOAEC. Although not explicitly stated, this approach is also applied for systemic endpoints.

In the “Guideline for quantifying substance-specific exposure-risk relationships and risk concentrations for exposure to carcinogenic substances at the workplace (AGS 2013) the BMD approach is mentioned as an alternative to the T25 approach. Special instructions regarding the application of the approach are given. However, it is noted that the instructions are not up to date compared the current recommendations by EFSA.

In the guidance on assessment factors to derive a DNEL published by ECETOC (2010) the benchmark approach is described as being preferred over the LOAEL-NAEL extrapolation. However ECETOC does not distinguish explicitly between continuous and quantal data. In case that only a LOAEL is available, ECETOC recommends using the $BMDL_{05}$ (presumably for continuous data) which they consider to correspond to the NOAEL.

In the basic scheme for the derivation of guide values for indoor air in Germany (no OEL values, values for the general population!) published as a notification in the “Bundesgesundheitsblatt” (Ad-hoc-AG 2012) the LOAEC is compared to the $BMDL_{10}$ and the NOAEC to the $BMDL_{05}$ (in case of epidemiological data a $BMDL_{01}$).

In general, the benchmark approach has found its way in the guidance documents for the derivation of OEL values. However, the information given in the documents is very limited; in some cases not even the selection of the BMR is indicated, often the BMD approach is only listed as an alternative to the LOAEL-NAEL extrapolation. This does not represent the description that this method represents the state of the science for determining a point of departure (POD) for risk assessment (Haber et al. 2018).

With the new versions of the benchmark tools currently available and the interest of the scientific community in this approach, the BMD approach should get more attention in the guidance documents and detailed instructions should be provided.

3.2 Practical recommendations

The BMD approach is a scientifically accepted and recommended way to derive a POD for risk assessment purposes. However, as shown in this report, there are decisions to be made by the risk assessor which recommendation to follow at several stages of the process.

As described in section 2.3.3 the new WHO IPCS document on the principles of dose-response modeling for risk assessment of chemicals (WHO 2020) gives advice for risk assessor how to perform a BMD analysis considering all currently available guidance and literature. Based on currently existing guidance the following recommendations can be given:

1) It is recommended to use one of the following tools:
   - EFSA web tool
   - BMDS

   These tools are developed for the specific purpose to be used by risk assessors. They were extensively tested in the past. The EFSA web tool is based on PROAST. It is well maintained and it was our observation, that it is updated always before PROAST and the PROAST web tool.

2) If guidance on specific aspects is required, it is recommended to consult the EFSA guidance (EFSA Scientific Committee et al. 2017). The guidance from the US EPA is from 2012 and does not represent the state of science from today.

3) Selection of BMR
   - Quantal data
     For quantal (animal) data selecting a BMR of 10% (extra risk) is generally accepted for experimental data. Other effect levels may be supported by statistical and biological characteristics of the data on a case-by-case basis. Other incidence levels might be applicable in case of human data with large cohorts.

   - Continuous data
     For continuous data the decision on the BMR should be driven by toxicological criteria. The BMR should mark the borderline to adversity. If no such criteria can be found, statistical criteria or defined changes compared to background levels might be the second choice.

4) Model averaging for both quantal and continuous data is the method of choice and renders model selection unnecessary.

5) As POD for further risk assessments the BMDL as calculated by the software based on model averaging is recommended (if not explicitly stated otherwise in the guidance used for the following assessment).

3.3 Usability of dose-response models to predict potency/risks above threshold

Dose-response modelling does not only provide an estimate of the POD for a defined benchmark response, but allows predicting expected responses at higher concentrations or doses.

Again, the type of information provided differs for different data types:
For continuous dose-response data the modelling will give back the estimated size of the continuous parameter at each selected dose or concentration. For example, if differences in body weight relative to the control is the critical endpoint and 10% difference is chosen as BMR to define the BMD and BMDL, then the relative difference to controls at various doses above the BMD can be calculated, using the same model (e.g. 17% at twice the BMD). In case of quantal data, the estimated incidence (as central estimate) at doses above the BMD can be calculated.

Note that in both cases only the effect levels under the conditions of the test and test system are calculated (for example, the mammalian in vivo test system used to generate the dose-response data). Interpretation of expected effect levels in the human target population (workers) is more difficult and will be exemplified:

Example:
In this example, derived from dose-response data identified in a subchronic inhalation study with rats, the BMD for a continuous effect (10% relative body weight decrease) is 10 mg/m^3. For simplicity it is assumed here that the BMD_{10} serves as the POD. With (arbitrarily assumed, for demonstration purposes only) assessment factors of 5 for intraspecies extrapolation, and both 3 for inter- and time extrapolation an OEL of 0.22 mg/m^3 is derived in this example.

A risk manager might be interested to get information on the expected effects in a worker population at a high long-term concentration. From the dose-response model it can be estimated that the relative body weight decrease under the experimental conditions is 24% at 70 mg/m^3. In this example the exposure conditions (duration of inhalation exposure, physical activity) are assumed to be representative for the workplace situation; therefore, no correction is required.

An intraspecies assessment factor is required to account for variability within the target population. An (arbitrary) factor of 5 is used here to account for the difference in susceptibility between the average worker and the more sensitive (represented as the 5th percentile of the distribution). In conclusion, at 14 mg/m^3 we would expect 24% relative body weight decrease in 5% of the target population.

However, additional assessment factors are typically applied to consider additional uncertainties, with regard to interspecies extrapolation and time extrapolation. They are applied to cover a chosen percentage of possible cases (chemicals). For the specific case they might be or might not be required, depending on a substance's properties in relation to the large set of chemicals forming the empirical background.

As used above for deriving the OEL, for both extrapolations (arbitrary) factors of 3 are applied.

The conclusion in this example is that a 24% relative body weight decrease in 5% of the target population can be expected for a concentration in the range between 14 mg/m^3 and 1.56 mg/m^3 (14 divided by 3 x 3) (see **Figure 3-1**).

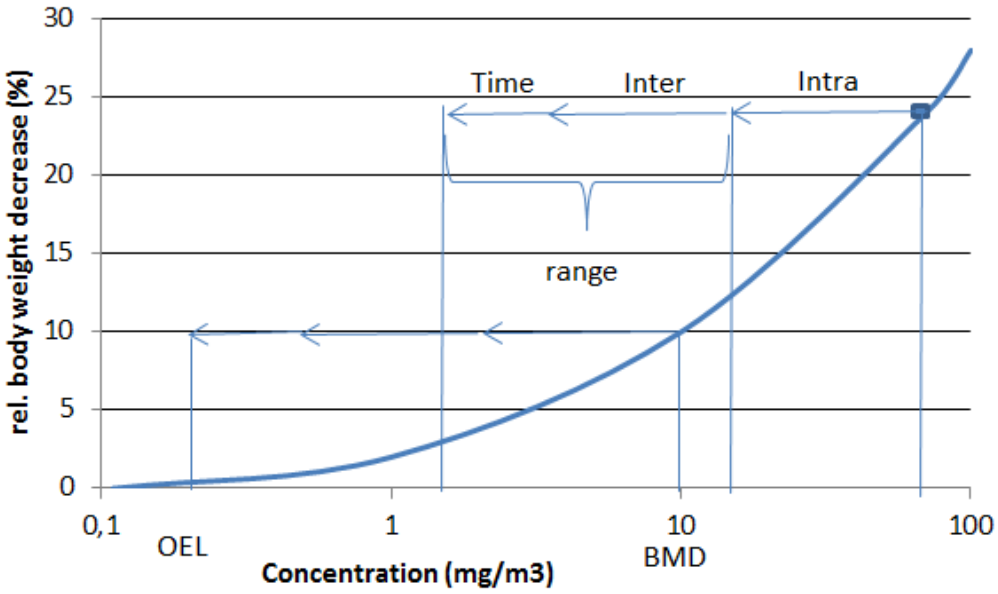


Figure 3-1 Exemplification of conclusions for effects occurring in the target population at concentrations above the OEL

The resulting concentration range is large and its interpretation difficult. A much better interpretation of effects above the OEL is possible with a probabilistic assessment, where the incidence of effects can be derived at various concentration levels, together with a description of the adhering uncertainty (this is discussed in more detail in a separate report on “Probabilistic assessments”).

4 Examples

Ten examples for demonstrating dose-response modelling practice are presented in a separate report.

References

Ad-hoc-AG. (2012). Richtwerte für die Innenraumluft: erste Fortschreibung des Basisschemas. Mitteilung der Ad-hoc-Arbeitsgruppe Innenraumrichtwerte der Innenraumlufthygiene-Kommission des Umweltbundesamtes und der Obersten Landesgesundheitsbehörden. Bundesgesundheitsblatt 55, 279-290.

AGS, Ausschuss für Gefahrstoffe. (2013). Leitfaden zur Quantifizierung stoffspezifischer Expositions-Risiko-Beziehungen und von Risikokonzentrationen bei Exposition gegenüber krebserzeugenden Gefahrstoffen am Arbeitsplatz, (Anlage 3 zu TRGS 910). Version N10, Stand: 15.09.2013, Ausarbeitung durch den Arbeitskreis Risikoableitung im UA III des AGS. http://www.baua.de/de/Themen-von-A-Z/Gefahrstoffe/TRGS/TRGS-910.html.

Allen, B.C., Kavlock, R.J., Kimmel, C.A., Faustman, E.M. (1994a). Dose-response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. Fundamental and Applied Toxicology 23, 487-495.

Allen, B.C., Kavlock, R.J., Kimmel, C.A., Faustman, E.M. (1994b). Dose-response assessment for developmental toxicity. III. Statistical models. Fundamental and Applied Toxicology 23, 496-509.

Barnes, D.G., Daston, G.P., Evans, J.S., Jarabek, A.M., Kavlock, R.J., Kimmel, C.A., Park, C., Spitzer, H.L. (1995). Benchmark dose workshop: criteria for use of a benchmark dose to estimate a reference dose. Regulatory Toxicology and Pharmacology 21, 296-306.

Bokkers, B.G.H., Slob, W. (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. Critical Reviews in Toxicology 37, 355-373.

Chen, C.C., Chen, J.J. (2014). Benchmark dose calculation for ordered categorical responses. Risk Analysis 34, 1435-1447.

Crump, K. (2002). Critical issues in benchmark calculations from continuous data. Critical Reviews in Toxicology 32, 133-153.

Crump, K.S. (1995). Calculation of benchmark doses from continuous data. Risk Analysis 15, 79-90.

Davis, J.A., Gift, J.S., Zhao, Q.J. (2011). Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1. Toxicology and Applied Pharmacology 254, 181-191.

Dekkers, S., de Heer, C., Rennen, M.A.J. (2001). Critical effect sizes in toxicological risk assessment: a comprehensive and critical evaluation. Environmental Toxicology and Pharmacology 10, 33-52.

DFG, Deutsche Forschungsgemeinschaft. (2019). MAK- und BAT-Werte-Liste 2019. Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe. Mitteilung 55, WILEY-VCH Verlag GmbH, Weinheim. https://onlinelibrary.wiley.com/doi/book/10.1002/9783527826155.

ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals. (2010). Technical Report No. 110. Guidance on Assessment Factors to Derive a DNEL, Brussels, Belgium.

ECHA, European Chemicals Agency. (2012). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1, November 2012, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

ECHA, European Chemicals Agency. (2018). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Draft Version 3.0. November 2018, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

ECHA, European Chemicals Agency. (2019). Guidance on Information Requirements and Chemical Safety Assessment. Appendix to Chapter R.8: Guidance for preparing a scientific report for health-based exposure limits at the workplace. Draft Public Version 1.0 May 2019, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

EFSA, European Food Safety Authority. (2009a). Guidance of the Scientific Committee on Use of the benchmark dose approach in risk assessment SC opinion on BMD. The EFSA Journal 1150, 1-72. http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2009.1150/epdf.

EFSA, European Food Safety Authority. (2009b). Scientific Opinion on Cadmium in Food, Scientific Opinion of the Panel on Contaminants in the Food Chain. The EFSA Journal 980, 1-139. https://www.efsa.europa.eu/de/efsajournal/pub/980.

EFSA, European Food Safety Authority. (2011). Use of BMDS and PROAST software packages by EFSA Scientific Panels and Units for applying the Benchmark Dose (BMD) approach in risk assessment. Technical Report EN113, Parma, Italy http://www.efsa.europa.eu/en/supporting/doc/113e.pdf.

EFSA, European Food Safety Authority. (2018). Manual for BMD Modeling (online application), https://r4eu.efsa.europa.eu/app/bmd.

EFSA Scientific Committee, Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, K.H., More, S., Mortensen, A., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Silano, V., Solecki, R., Turck, D., Aerts, M., Bodin, L., Davis,

A., Edler, L., Gundert-Remy, U., Sand, S., Slob, W., Bottex, B., Abrahantes, J.C., Marques, D.C., Kass, G., Schlatter, J.R. (2017). Update: use of the benchmark dose approach in risk assessment. The EFSA Journal 15(1):4658www, 41pp.

Fang, Q., Piegorsch, W.W., Simmons, S.J., Li, X., Chen, C., Wang, Y. (2015). Bayesian model-averaged benchmark dose analysis via reparameterized quantal-response models. Biometrics 71, 1168-1175.

Faustman, E.M., Allen, B.C., Kavlock, R.J., Kimmel, C.A. (1994). Dose-response assessment for developmental toxicity. I. Characterization of database and determination of no observed adverse effect levels. Fundamental and Applied Toxicology 23, 478-486.

Gaylor, D.W., Slikker, W. (1990). Risk assessment for neurotoxic effects. Neurotoxicology 11, 211-218.

Haber, L.T., Dourson, M.L., Allen, B.C., Hertzberg, R.C., Parker, A., Vincent, M.J., Maier, A., Boobis, A.R. (2018). Benchmark dose (BMD) modeling: current practice, issues, and challenges. Critical Reviews in Toxicology 48, 387-415.

Kalantari, F., Ringblom, J., Sand, S., Öberg, M. (2017). Influence of distribution of animals between dose groups on estimated benchmark dose and animal distress for quantal responses. Risk Analysis 37, 1716-1728.

Kavlock, R.J., Allen, B.C., Kimmel, C.A., Faustman, E.M. (1995). Dose-response assessment for developmental toxicity. IV. Benchmark doses for fetal weight changes. Fundamental and Applied Toxicology 26, 211-222.

Kodell, R. (2009). Replace the NOAEL and LOAEL with the $BMDL_{01}$ and $BMDL_{10}$. Environmental and Ecological Statistics 16, 3-12.

Kwon, D., Hoffman, F.O., Moroz, B.E., Simon, S.L. (2016). Bayesian dose-response analysis for epidemiological studies with complex uncertainty in dose estimation. Statistics in Medicine 35, 399-423.

Muri, S.D., Schlatter, J.R., Brüschweiler, B.J. (2009). The benchmark dose approach in food risk assessment: is it applicable and worthwhile? Food and Chemical Toxicology 47, 2906-2925.

RIVM, Rijksinstituut voor Volksgezondheid en Milieu. (2019). PROAST Manual. Menu version, PROAST verions 65.6. 6 February 2019.

Sand, S., Portier, C.J., Krewski, D. (2011). A signal-to-noise crossover dose as the point of departure for health risk assessment. Environmental Health Perspectives 119, 1766-1774.

Sand, S., Victorin, K., Falk Filipsson, A. (2008). The current state of knowledge on the use of the benchmark dose concept in risk assessment. Journal of Applied Toxicology 28, 405-421.

Sand, S., von Rosen, D., Victorin, K., Falk Filipsson, A. (2006). Identification of a critical dose level for risk assessment: developments in benchmark dose analysis of continuous endpoints. Toxicological Sciences 90, 241-251.

Schneider, K., Kaiser, E. (2012). Anwendung des Benchmark-Verfahrens bei der Ableitung von HBM-Werten [Application of the benchmark approach for derivation of HBM values (human biomonitoring values)], Forschungs- und Beratungsinstitut Gefahrstoffe GmbH (FoBiG), Freiburg. Under contract for the German Federal Environmental Agency (Umweltbundesamt) Berlin/Dessau, FKZ 363 01 383.

SCOEL, Scientific Committee for Occupational Exposure Limits. (2013). Methodology for the Derivation of Occupational Exposure Limits. Key Documentation (version 7), June 2013, European Commission, Employment, Social Affairs and Inclusion.

Shao, K. (2012). A comparison of three methods for integrating historical information for Bayesian model averaged benchmark dose estimation. Environmental Toxicology and Pharmacology 34, 288-296.

Shao, K., Gift, J.S. (2014). Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. Risk Analysis 34, 101-120.

Shao, K., Shapiro, A.J. (2018). A Web-Based System for Bayesian Benchmark Dose Estimation. Environmental Health Perspectives 126, 017002.

Slob, W. (2014a). Benchmark dose and the three Rs. Part I. Getting more information from the same number of animals. Critical Reviews in Toxicology 44, 557-567.

Slob, W. (2014b). Benchmark dose and the three Rs. Part II. Consequences for study design and animal use. Critical Reviews in Toxicology 44, 568-580.

Slob, W. (2017). A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints. Critical Reviews in Toxicology 47, 342-351.

Slob, W. (2018). Joint project on Benchmark Dose modelling with RIVM. EFSA supporting publication 2018:EN-1497, 14 pp. doi:10.2903/sp.efsa.2018.EN-1497.

US EPA, Environmental Protection Agency. (2012). Benchmark Dose Technical Guidance. EPA/100/R-12/001, Risk Assessment Forum. U.S. Environmental Protection Agency, Washington, DC. https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf.

US EPA, Environmental Protection Agency. (2018). Benchmark Dose Software (BMDS). Version 3.0. User Guide. EPA/600/R-18/310,

https://www.epa.gov/bmds/benchmark-dose-software-bmds-version-30-user-guide-readme.

Varewyck, M., Verbeke, T. (2017). Software for benchmark dose modelling. EFSA Supporting Publications 14, 15pp.

Wheeler, M., Bailer, A. (2009). Comparing model averaging with other model selection strategies for benchmark dose estimation. Environmental and Ecological Statistics 16, 37-51.

Wheeler, M.W., Bailer, A.J. (2007). Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Analysis 27, 659-670.

Wheeler, M.W., Bailer, A.J. (2008). Model averaging software dor dichotomous dose response risk estimation. Journal of Statistical Software 26, 1-15.

WHO, World Health Organization. (2014). Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. IPCS Harmonization Project Document 11, WHO, Geneva, Switzerland. http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf.

WHO, World Health Organization. (2020). Environmental Health Criteria 240, Principles and Methods for the Risk Assessment of Chemicals in Food. Chapter 5: Dose–response assessment and derivation of health-based guidance values, IPCS International Programme on Chemical Safety; World Health Organization, Geneva. https://www.who.int/docs/default-source/food-safety/publications/chapter5-dose-response.pdf?sfvrsn=32edc2c6_5.

Wignall, J.A., Shapiro, A.J., Wright, F.A., Woodruff, T.J., Chiu, W.A., Guyton, K.Z., Rusyn, I. (2014). Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. Environmental Health Perspectives 122, 499-505.

Zeller, A., Duran-Pacheco, G., Guérard, M. (2017). An appraisal of critical effect sizes for the benchmark dose approach to assess dose-response relationships in genetic toxicology. Archives of Toxicology 91, 3799-3807.

List of Tables

List of Figures

REPORT 3:
Benchmark Dose Modelling - Examples

RESEARCH PROJECT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Abbreviations

3-MCPD	3-Monochloropropane-1,2-diol
AGS	Ausschuss für Gefahrstoffe
AIC	Akaike information criterion
BMD	Benchmark dose
BMDL	Benchmark dose lower bound
BMDU	Benchmark dose upper bound
BMR	Benchmark response
BMDS	Benchmark dose software
EFSA	European Food and Safety Authority
MOCA	4,4’-Methylene-bis-[2-chloroaniline]
LOAEL	Lowest observed adverse effect level
MAK	Maximale Arbeitsplatzkonzentration
NAEL	No adverse effect level
NOAEC	No observed adverse effect concentration
NOAEL	No observed adverse effect level
NTA	Nitrilotriacetic acid
OBPP	N-octadecyl β-(3',5'-di-tert-butyl-4'-hydroxyphenyl)propionate
OEL	Occupational exposure limit

POD	Point of departure
PROAST	Dose-response modelling software by RIVM

1 Introduction

1.1 Example substances

In this report benchmark dose (BMD) modelling for ten example substances is presented.

Substances and the corresponding data were selected based on the following criteria:
- Five substances with quantal data and five substances with continuous data,
- datasets with NOAELs and datasets where only LOAELs could be determined,
- one dataset with epidemiological data.

The following table gives an overview of the substances and datasets selected for benchmark dose modelling in this report.

Table 1-1 Overview of the substances, studies and endpoints selected for benchmark modelling in this report

Substance	**Dataset**
Quantal data (using EFSA PROAST 67.0)	
3-Monochloropropane-1,2-diol (3-MCPD)	Cho et al. (2008) Study type: chronic toxicity study in rats Effect: renal tubule hyperplasia in males
Divanadium pentaoxide	NTP (2002), NTP TR No. 507 Study type: chronic toxicity study in rats Effect: chronic inflammation in the lung of males and females
4,4'-Methylene-bis-[2-chloroaniline] (MOCA)	Kommineni et al. (1979) and RAC (2017) Study type: chronic toxicity study in rats Effect: lung tumours (adenomas, epidermoid carcinomas, adenocarcinomas)
Nitrilotriacetic acid	Greim and MAK Kommission (2008) Study type**:** chronic toxicity study in rats Effect: hyperplasia of the transitional epithelium of the urinary bladder
Benzoic acid	Hartwig and MAK Commission (2018) Study type: 28-day study in rats

Substance	**Dataset**
	Effect: Interstitial infiltration of inflammatory cells in the lung (generalized effect)
Continuous data (using EFSA PROAST 69.0)	
Nalidixic acid	NTP (1989), NTP TR No. 368 Study type: chronic toxicity study in rats Effect: body weight changes in male and female rats
1,1,2,2 Tetrachloroethane	NTP (2004), NTP Toxicity report No. 49 Study type: 14-week study in rats Effect: changes in relative liver weight and sperm motility
N-octadecyl β-(3',5'-di-tert-butyl-4'-hydroxyphenyl)propionate (OBPP)	Lake et al. (1980) Study type: 14 days study in rats Effect: changes in relative liver weight
Tert-Butyl alcohol	NTP (1995), NTP TR No. 436, Hartwig (2014) Study type: chronic toxicity study in rats Effect: increase in relative kidney weight in females
Benzene	Zhang et al. (2016) Study type: epidemiological data Effect: reduced white blood cell count in workers

1.2 Basic information on benchmark modelling

As pointed out in the report “Benchmark dose modelling” of this research project, benchmark dose modelling can be performed with different statistical tools which are available online. It was decided to use the PROAST-based web tools from EFSA[1] for modelling all datasets in the current report since it corresponds to the EFSA guidance for BMD modelling, is regularly updated and a report of the modelling results in “doxc-format” can easily be generated. For quantal data, modelling was performed with the EFSA tool under PROAST version 67.0, modelling of continuous data was done after the tool was updated to PROAST version 69.0 in June 2020. For both data types model averaging was applied.

[1] https://r4eu.efsa.europa.eu/

7 R3: Benchmark dose modelling - examples

For two substances (3-MCPD and benzoic acid) modelling was also performed with the BMDS software from the US EPA[2]. For 3-MCPD additional modelling was performed with the PROAST web tool[3], and the PROAST versions implemented in the R software[4] (GUI and MENU version).

Details and background information on benchmark dose modelling is provided in the separate report on “Benchmark dose modelling”. The following section shortly summarises the main descriptors used in benchmark dose modelling to give an understanding of the data and results presented in section 2.

The benchmark dose (BMD) describes a dose with a predefined additional effect compared to the control. The effect size related to the BMD is called benchmark response (BMR). Confidence intervals express the uncertainty due to sampling and/or measurement error. The benchmark dose lower bound (BMDL) refers to the lower limit of a (in general) 95th confidence interval on the BMD, the BMDU (benchmark dose upper limit) refers to the upper limit.

According to WHO (2009) and the updated WHO IPCS “Chapter 5. Dose-response Assessment and Derivation of health-based Guidance values” (WHO 2020), a benchmark response (BMR) of 10% given as extra risk was selected for all quantal datasets modelled here. Extra risk is defined as an absolute change in frequency of response divided by the non-affected fraction in the control population (100 minus the background response in %) (EFSA Scientific Committee et al. 2017). For continuous data, the BMR was selected for each substance and dataset individually and was based on toxicological criteria. This means that a response above the BMR was considered as adverse. For details see the examples in the next section.

To account for the uncertainty of each single model, model averaging (averaging the individual model results based on their goodness of fit) is performed by the modelling tools. Better fitting models weighs more in model averaging than a model which fits less good to the data. In the end, the BMD and BMDL are calculated based on the “average model”. Only BMDL and BMDU, but no “average” BMD is calculated in the PROAST-based web tools.

[2] https://www.epa.gov/bmds/benchmark-dose-software-bmds-version-312-download
[3] https://proastweb.rivm.nl/
[4] https://www.rivm.nl/en/media/89001

2 Examples

2.1 Quantal data

2.1.1 3-Monochloropropane-1,2-diol (3-MCPD)

In the chronic toxicity study of Cho et al. (2008) renal tubule hyperplasia in rats was reported (see the following table). As laid out by Haber et al. (2018) (see below) this endpoint was considered relevant and selected by EFSA and JECFA for the derivation of TDI values for the substance.

Table 2-1 Data on renal tubule hyperplasia in male rats (according to Cho et al. (2008)) used for benchmark modelling

dose (mg/kg bw/day)	effect # affected animals	n # animals in group
0.00	1	50
1.97	11	50
8.27	21	50
29.50	36	50

Modelling parameters:

- BMR: 10% extra risk
- Model averaging

Result:

The complete report generated with the EFSA web tool is included in the Annex.

BMDL: 0.19 mg/kg bw/d
BMDU: 1.88 mg/kg bw/d

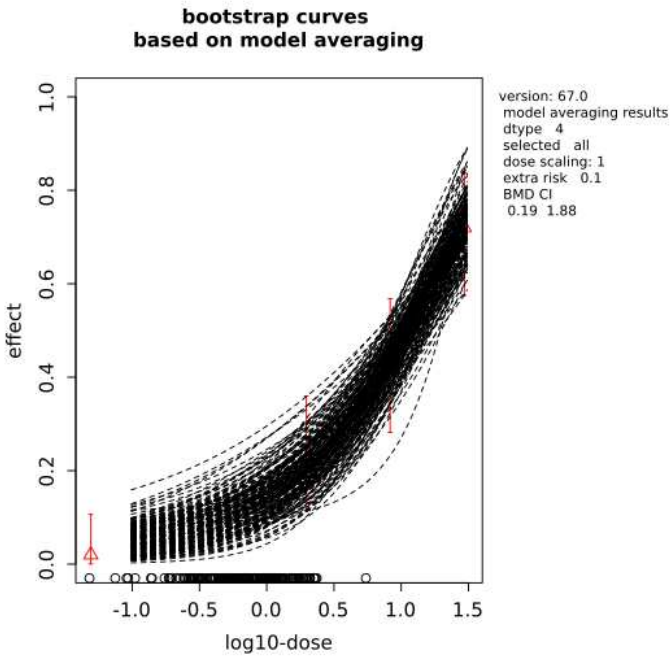


Figure 2-1 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool).

In addition to modelling the data with the EFSA web tool, modelling was performed with the PROAST web tool, BMDS 3.1 and PROAST v67.0 running in "R". For results see the following table.

Table 2-2 Comparison of results (BMD, BMDL and BMDU) from benchmark dose modelling with different tools.

Tool	Method	BMD	BMDL	BMDU
EFSA-tool	Averaging	-	0.19	1.88
PROAST-web	Averaging	-	0.193	1.88
BMDS 3.1	Bayesian Averaging	1.87	0.62	3.11
PROAST v67.0 (GUI version)	Single model	-	0.074 (lowest BMDL, gamma)	1.93 (highest BMDU, log probit)
PROAST v67.0 (MENU version)	Averaging	-	0.193	1.880

Discussion / Comparison with NOAEL:
Until about mid of June 2020 the two PROAST-based web tools and PROAST in “R” used the same version number of PROAST (v67.0). This is reflected in the identical results after model averaging. On 19.6.2020 the EFSA web tool was updated to PROAST version 69.0, which included for example a bug fix for model averaging of continuous data.

When using the PROAST GUI version which does not provide model averaging, the BMDL and the BMDU presented here are the lowest/highest values from all accepted models. BMDS also applies model averaging, however this tool uses a different averaging strategy (Bayesian averaging). This may also be a reason for different results compared to the PROAST-based tools.

The following table shows a comparison of the NOAEL derived by the authors of this document and the BMDL calculated with the EFSA-tool.

Table 2-3 Comparison of NOAEL and BMDL

NOAEL (derived by authors of this document)	LOAEL (derived by authors of this report)	BMDL
-*	1.97 mg/kg bw/d	0.19 mg/kg bw/d

*NOAEL cannot be determined since significant effects (>20% incidence) were already observed at the lowest dose tested

According to Haber et al. (2018) the results presented in the following table were obtained for the same endpoint (and in most cases also the same dataset). For the original references please see Haber et al. (2018).

Table 2-4 BMD-modelling results for MCPD reported according to Haber et al. (2018)

	BMD (mg/kg bw/d)	**BMDL (mg/kg bw/d)**	**Dataset**	**BMD model**
Haber et al. (2018)	1.2	0.87	Cho et al. (2008), male only	Log-logistic (restricted)
Haber et al. (2018)	1.5	0.74	Cho et al. (2008), male only	Average of 9 models
Abraham et al (2012)	0.92	0.27	Cho et al. (2008), male only	Log-probit (unrestricted)
EFSA (2016)	0.54	0.077	Cho et al. (2008), male only	Gamma (unrestricted)
EFSA (2018)	0.68	0.20	Cho et al. (2008), male only	Model averaging
JECFA (2016, 2017)	1.2	0.87	Cho et al. (2008), male only	Log-logistic (restricted)
JECFA (2016, 2017)	1.29	0.89	Cho et al. (2008), male only	Model averaging
Rietjens et al. (2012)	1.27	0.72	Cho et al. (2008) and Sunahara (1993)	Average of 7 models
Hwang et al. (2009)	1.2	0.87	Cho et al. (2008), male only	Log-logistic (restricted)
This evaluation	**-**	**0.19**	**Cho et al. (2008), male only**	**Model averaging**

The BMDL determined in the current evaluation (0.19 mg/kg bw/d) matches with the evaluation from EFSA in 2018 (BMDL: 0.20 mg/kg bw/d) which also used the same dataset and model averaging. The differences to the other evaluations may be explained by using only one model (no averaging), older versions of the BMD modelling software or the use of combined datasets from several studies.

2.1.3 Divanadium pentaoxide

In the chronic toxicity study of NTP (2002) chronic inflammation of the lung in male and female rats was observed after inhalation exposure (see the following table).

Table 2-5 Data on chronic inflammation of the lung in male and female rats (according to NTP (2002)) used for benchmark dose modelling

concentration (mg/m³)	effect # affected animals	n # animals in group	sex
0	10	49	f
0.28	10	49	f
0.56	14	50	f
1.12	40	50	f
0	5	50	m
0.28	8	49	m
0.56	24	48	m
1.12	42	50	m

m, f = male, female

Modelling parameters:

- BMR: 10% extra risk
- Model averaging
- Data from both sexes combined in one dataset (sex not considered as a covariate)

Result:
The complete report generated with the EFSA web tool is included in the Annex.

BMDL: 0.23 mg/m³
BMDU: 0.46 mg/m³

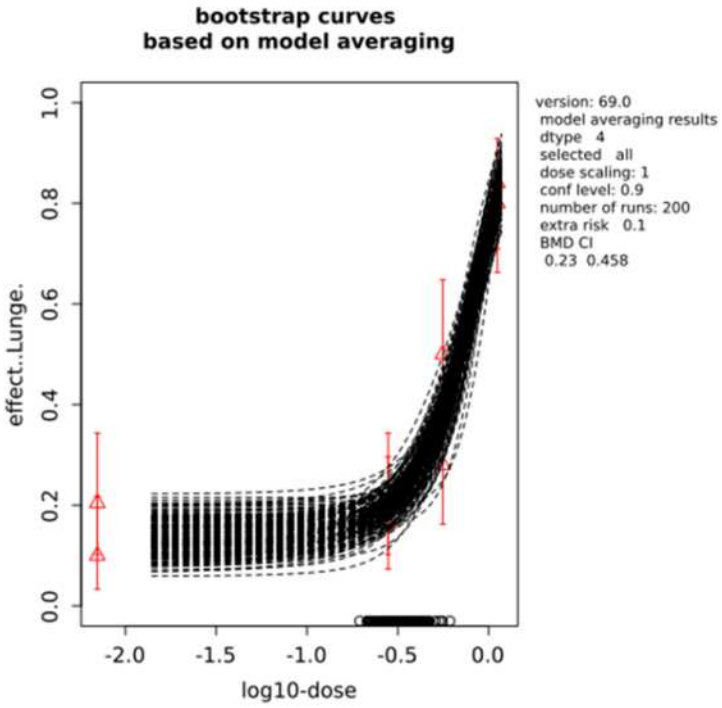


Figure 2-2 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool)

Discussion / Comparison with NOAEL:
The following table shows a comparison of the NOAEL derived by the authors of this document and the BMDL calculated with the EFSA-tool.

Table 2-6 Comparison of NOAEC and BMDL

NOAEC (derived by authors of this document)	LOAEC (derived by authors of this report)	BMDL
-*	0.28 mg/m³	0.23 mg/m³

*NOAEL cannot be determined since effects were already observed in male and female animals at the lowest dose tested

2.1.4 4,4'-Methylene-bis-[2-chloroaniline] (MOCA)

In the chronic toxicity study (18 months) from Kommineni et al. (1979) lung tumours (adenomas, epidermoid carcinomas, adenocarcinomas) were observed in male rats. The dose and incidences for benchmark modelling were used as given by (RAC 2017) (see the following table).

Table 2-7 Data on lung tumours in male rats (according to Kommineni et al. (1979) and (RAC 2017) used for benchmark dose modelling

Corrected dose (mg/kg bw/d)	effect # affected animals	n # animals in group
0	1	100
9.4	23	100
18.8	28	75
37.5	35	50

Modelling parameters:

- BMR: 10% extra risk
- Model averaging

Result:
The complete report generated with the EFSA web tool is included in the Annex.

BMDL: 2.91 mg/kg bw/d
BMDU: 7.19 mg/kg bw/d

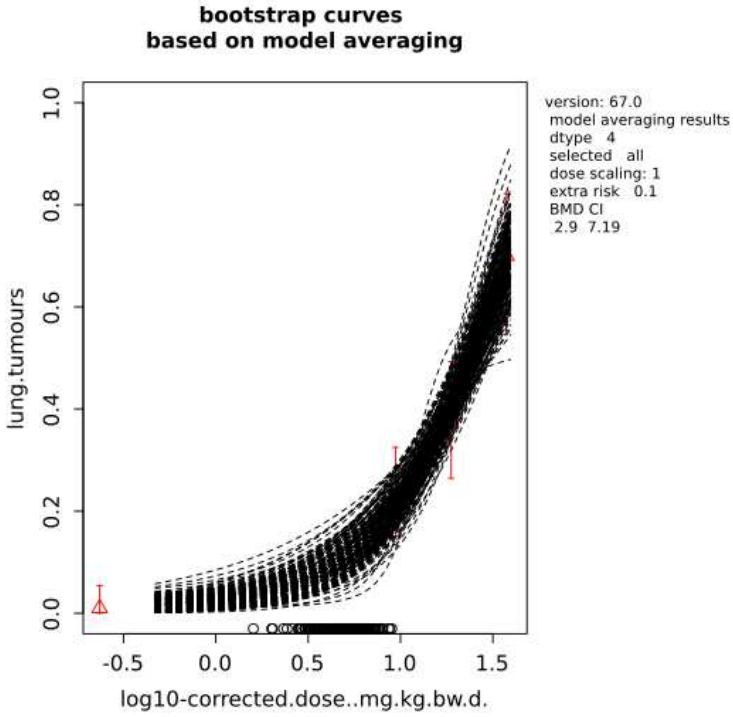


Figure 2-3 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool)

Discussion / Comparison with NOAEL:
The following table shows a comparison of the T25 (oral, rat) calculated by RAC (2017) based on the same study and the $BMDL_{10}$ calculated with the EFSA-tool. Considering the different "response levels" (25% incidence for T25 and 10% incidence for BMDL), the values show good agreement.

Table 2-8 Comparison of T25 derived by RAC (2017) and BMDL

T25 (oral, rats)	$BMDL_{10}$
10.6 mg/kg bw/d	2.91 mg/kg bw/d

2.1.5 Nitrilotriacetic acid (NTA) and its sodium salts

In the chronic toxicity study by NCI from 1977 reported in the MAK documentation (Greim 2008) hyperplasia of the transitional epithelium of the urinary bladder in male and female rats was observed after exposure to trisodium nitrilotriacetate (Na_3NTA, see the following table).

Table 2-9 Data on urinary bladder transitional epithelium hyperplasia in male and female rats (according to MAK Commission) used for benchmark dose modelling (exposure to trisodium nitrilotriacetate)

Corrected dose (mg/kg bw/d)	effect # affected animals	n # animals in group	sex
0	0	24	m
10	3	23	m
100	3	24	m
1000	8	24	m
0	1	24	f
10	1	24	f
100	13	24	f
1000	14	24	f

m, f = male, female

Modelling parameters:

- BMR: 10% extra risk
- Model averaging
- Data from both sexes combined in one dataset (sex not considered as a covariate)

Result:
The complete report generated with the EFSA web tool is included in the Annex.

BMDL: 1.46 mg/kg bw/d
BMDU: 72.4 mg/kg bw/d

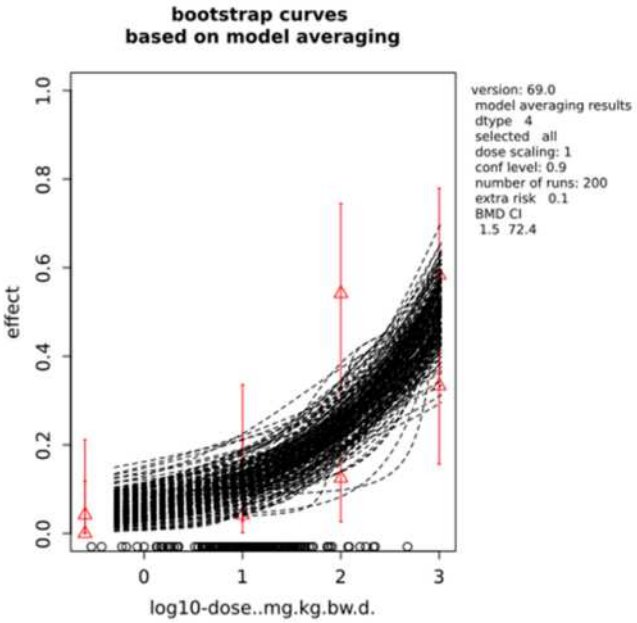


Figure 2-4 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool).

Discussion / Comparison with NOAEL:
The following table shows a comparison of the NOAEL derived by the MAK Commission[5] (not yet published) and the BMDL calculated with the EFSA-tool.

Table 2-10 Comparison of NAEL and BMDL for trisodium nitrilotriacetate and the calculated values for nitriloacetic acid

NAEL (derived by MAK Commission)	LOAEL (derived MAK Commission)	BMDL
LOAEL/3 = 2.3 mg NTA/kg bw/d	10 mg Na_3NTA/ kg bw/d (= 6.9 mg NTA/ kg bw/d)	1.46 mg Na_3NTA /kg bw/d (= 1.0 mg NTA/kg bw/d)

[5] Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area (MAK Commission) of the Deutsche Forschungsgemeinschaft DFG

2.1.7 Benzoic acid

In the 4-week inhalation toxicity study by Velsicol Chemical Company from 1981 reported in the MAK documentation for benzoic acid and alkali benzoates (Hartwig and MAK Commission 2018) interstitial inflammation and fibrosis of the lung in male and female rats was reported. For each concentration group and for both effects data are presented for “focal”, “multifocal” or “generalized” occurrences, representing an increase in severity in the presented order. Data on interstitial inflammation were selected for BMD-modelling and only “generalized” effects (not those classified as “focal” or “multifocal”) were transferred to a quantal dataset which was then modelled (see the following table).

Table 2-11 Data on interstitial inflammation (“generalized”) of the lung in male and female rats (according to MAK Commission) used for benchmark dose modelling

concentration (mg/m³)	effect # affected animals	n # animals in group	sex
0	0	10	m
25	3	10	m
250	4	10	m
1200	8	10	m
0	0	10	f
25	0	10	f
250	5	10	f
1200	9	10	f

m, f = male, female

Modelling parameters:

- BMR: 10% extra risk
- Model averaging
- Data from both sexes combined in one dataset (sex not considered as a covariate)

Result:
The complete report generated with the EFSA web tool is included in the Annex.

BMDL: 6.36 mg/m³
BMDU: 92 mg/m³

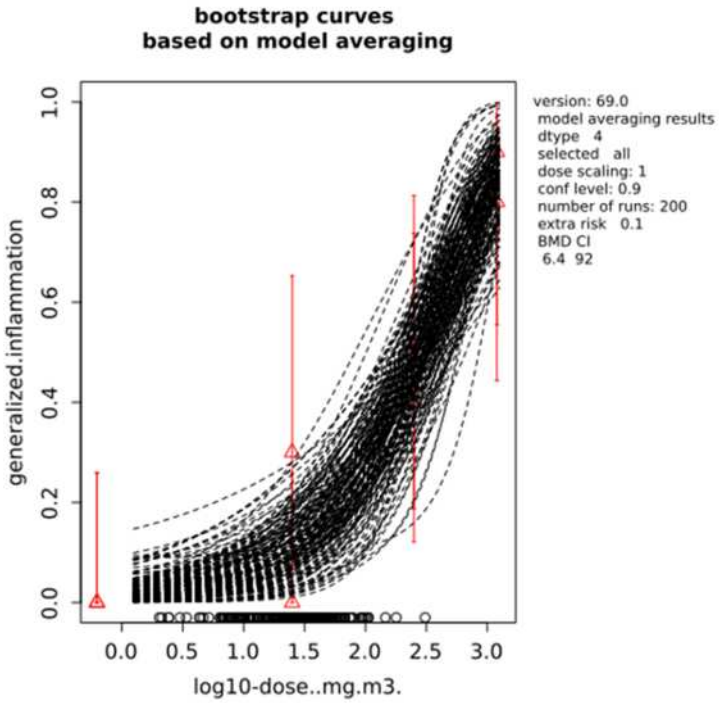


Figure 2-5 Graphical representation of the modelling results for male and female data combined (taken from the EFSA report generated with the web tool).

For benzoic acid modelling was also performed with BMDS 3.1 for the combined dataset, selecting the same BMR (10% extra risk) and model averaging.

BMD:	69.31 mg/m³
BMDL:	27.90 mg/m³
BMDU:	145.17 mg/m³

Discussion / Comparison with NOAEC:
The differences observed with the different modelling tools (EFSA web tool and BMDS 3.1) can be explained by the uncertainty of the data (visible in form of the vertical red lines in **Figure 2-5**): The lowest dose with effects for females and males differs (25 mg/kg/d for males and a 10fold higher dose for females), but both are associated with a high incidence (30% and 50%, respectively). So, data for males and females are diverging in the low dose range and for the relevant effect range (around 10%) data points are lacking. In consequence, dose-response data allow various shapes of the models used. In addition, the tools apply a slightly different set of models. In BMDS 3.1 the quantal linear model is given a posterior probability of 0.622, followed by the multistage model with 0.16. All the other models have a posterior probability of 0.01 – 0.06. In the EFSA web tool the Weibull model and gamma model are given most weigh in the averaging process. This, as described above, is only possible due to the high variability in the data.

The following table shows a comparison of the NOAEC derived by the MAK Commission (Hartwig and MAK Commission 2018) and the BMDL calculated with the EFSA-tool.

Table 2-12 Comparison of NOAEC and BMDL

NOAEC (according to MAK)	BMDL
12.6 mg/m³ (highest concentration tested in a 4-week inhalation study; no effects observed)	6.36 mg/m³

2.2 Continuous data

2.2.1 Nalidixic acid

In the chronic toxicity study from NTP (1989) effects on body weight changes in male and female rats are reported (see the following table).

Table 2-13 Data on body weight change in male and female rats (according to NTP (1989)) used for benchmark dose modelling

concentration in food (ppm)	Body weight (mean in g)	Body weight (SEM in g)	n # animals in group	sex
0	29.4	1.13	10	f
1000	28.2	1.18	10	f
2000	28.7	1.06	9	f
4000	27.1	0.42	10	f
8000	24.8	0.84	10	f
16000	23.6	0.55	10	f
0	36.1	0.89	10	m
1000	35.0	0.64	10	m
2000	34.9	0.71	10	m
4000	33.6	0.41	10	m
8000	32.4	0.47	10	m
16000	31.4	0.71	10	m

m, f = male, female

Modelling parameters:

- BMR: 10% difference in final body weight compared to the controls. According to Dekkers et al. (2001) BMRs of 5% or 10% change in final body weight are recommended for this endpoint (based on information including biological and

toxicological knowledge). Due to common practice, 10% change in final body weight were defined as the border to adversity.

- Model averaging
- Modelling performed for both sexes separately[6]

Result:
The complete reports generated with the EFSA web tool is included in the Annex.

BMDL female: 1650 ppm food
BMDU female: 10300 ppm food

BMDL male: 4410 ppm food
BMDU male: 14800 ppm food

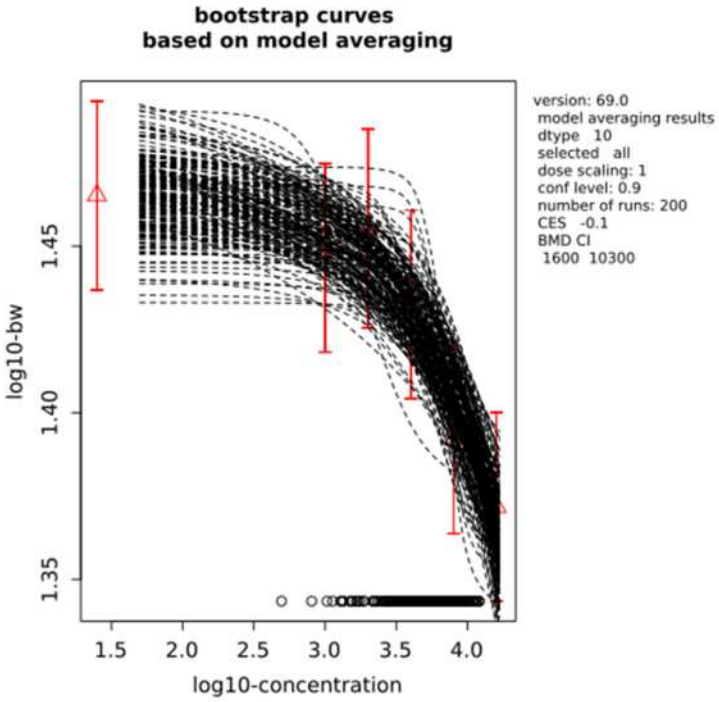


Figure 2-6 Graphical representation of the modelling results for female animals (taken from the EFSA report generated with the web tool)

[6] Data for both sexes were modelled separately since a combination of both datasets resulted in an "AIC warning". However, BMD-modelling of the combined dataset led to a BMDL of 1790 ppm food and a BMDU of 10300 ppm food.

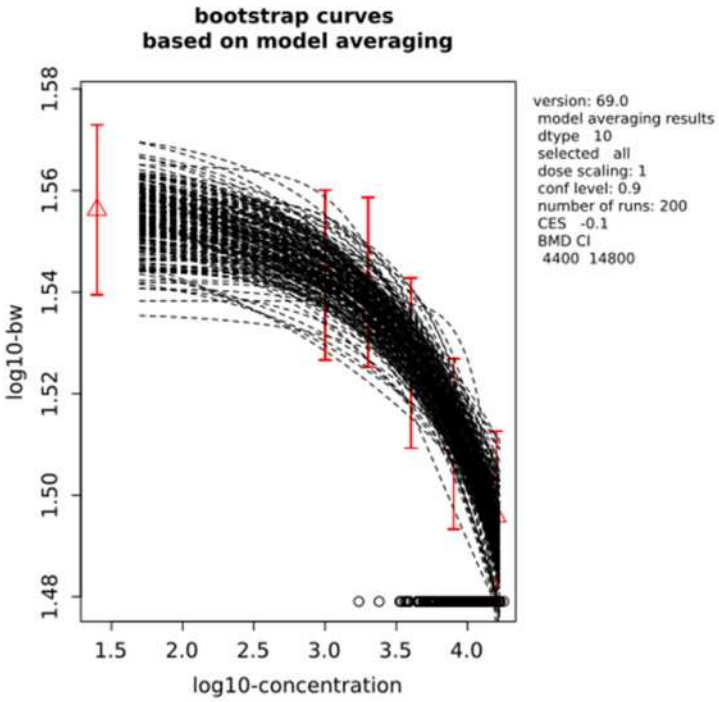


Figure 2-7 Graphical representation of the modelling results for male animals (taken from the EFSA report generated with the web tool)

Discussion / Comparison with NOAEL:
The following table shows a comparison of the NOAEL derived by the authors of this document and the BMDL for female rats calculated with the EFSA-tool.

Table 2-14 Comparison of NOAEL and BMDL

NOAEL (derived by authors of this document)	BMDL
4000 ppm food	1650 ppm food

2.2.2 1,1,2,2 Tetrachloroethane

In the 14-week feeding toxicity study from NTP (2004) an increase in relative liver weight (liver weight to body weight ratio) in male and female rats is reported (see

Table **2-15**). In addition, a decrease in sperm motility was observed in male animals (data are presented in
Table **2-16**). Sperm motility was only monitored in in the control and at 40, 80 and 170 mg/kg bw/d.

Table 2-15 Data on relative liver weight in male and female rats (according to NTP (2004) used for benchmark dose modelling

dose (mg/kg bw/d)	Relative liver weight (mean in mg organ weight / g bw)	Relative liver weight (SEM in mg organ weight / g bw)	n # animals in group	sex
0	34.79	0.42	10	m
20	36.72	0.44	10	m
40	41.03	0.85	10	m
80	45.61	0.52	10	m
170	44.68	0.45	10	m
320	52.23	1.42	10	m
0	35.07	0.56	10	f
20	36.69	0.36	10	f
40	37.84	0.51	10	f
80	44.2	0.27	10	f
170	48.03	0.89	10	f
320	58.4	1.42	10	f

m, f = male, female

Table 2-16 Data on sperm motility in male rats (according to NTP (2004)) used for benchmark modelling

dose (mg/kg bw/d)	Sperm motility (mean in %)	Sperm motility (SEM in %)	n # animals in group	sex
0	83.58	0.86	10	m
40	69.3	3.34	10	m
80	71.09	1.7	10	m
170	63.49	3.65	10	m

m = male

Modelling parameters:

- Selection of BMR

  A) BMR 7% for changes in relative liver weight

  Selection of BMR for changes in relative liver weight: According to Dekkers et al. (2001) a BMR of 5% based on human data is recommended for the ratio of

liver weight to body weight. Liver toxicity is a relevant endpoint. At higher doses hypertrophy and necrosis of the liver were observed.
A change of 5% from the control mean value for male animals would result in a hypothetical relative liver weight of 36.52 g for the onset of toxicologically relevant effects. This value is very close to the variability in the control group (34.79 + 1.32 = 36.11). Therefore, it was decided not to use a BMR of 5% but to calculate a BMR based on 2 standard deviations (SD) (to be sure that effects are not within the natural variance of the control group).
With a BMR of 2 SD from the control the following values were calculated:
For male rats: 34.79 (100%) + 2 SD = 37.446 (107.63%)
For female rats: 35.07 (100%) + 2 SD = 38.61 (110.09%)

With an overall BMR of 7% the onset of changes that can lead to the more severe liver effects should be covered.
Overall, a BMR of 7% was selected for modelling of both sexes.

B) BMR 20% for changes in relative liver weight

According to a common procedure agreed upon by the AGS, changes in relative liver weight >=20% are considered as adverse (following a conclusion of the MAK commission for defining the borderline between pure adaptive responses and beginning liver toxicity (unpublished minutes of Sub-Committee III of AGS, June 2016)). Therefore, an additional benchmark modelling with a BMR of 20% was performed.

- Selection of BMR for reduced sperm motility: Dekkers et al. (2001) do not give any recommendations for the selection of a BMR for adverse effects on sperm motility. Due to the very low SD in the control group (2.71, SEM transferred to SD)), the BMR was also set to control mean + 2 SD:

  For male rats: 83.58 (100%) - 2 SD = 78.142 (93.494%)

  A BMR of 6.5% was selected for modelling.

- Model averaging

Results:
The complete report (for “relative liver weight” with the combined dataset) generated with the EFSA web tool is included in the Annex.

Relative liver weight
Both modelling approaches (BMR 7% and 20%) resulted in a warning that “*the AIC of the best model (minimum AIC) is more than two units larger than that of the full model. This might indicate a problem in the data, in particular when the difference is much larger than two units (e.g. > 5)*”. Therefore, the data were remodelled using “sex” as a covariate. Nevertheless, the “AIC warning” also appeared for this modelling with the lowest BMDL for females being in the same range as the BMDL reported below for the

combined dataset. As a result, the BMDL for the modelling with the combined data set was selected.

A) Combined dataset (for males and females, BMR 7%):

BMDL: 7.87 mg/kg bw/d
BMDU: 22.9 mg/kg bw/d

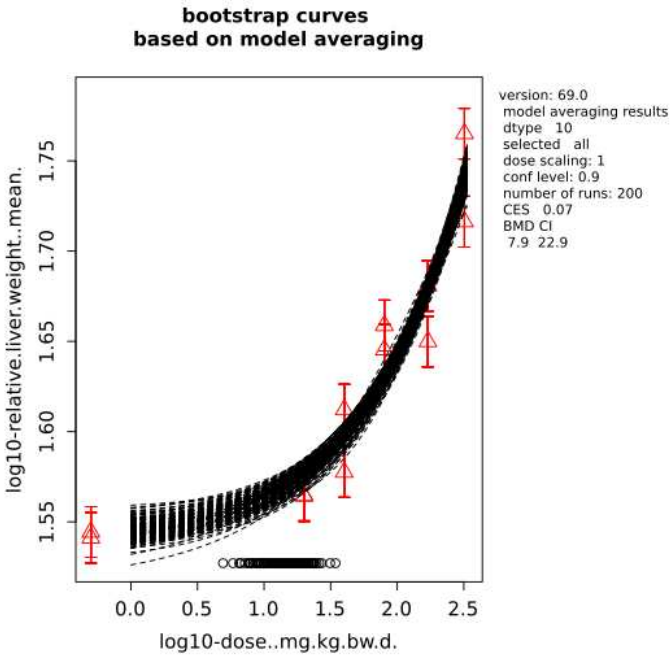


Figure 2-8 Graphical representation of the modelling results for the combined dataset (BMR 7%, taken from the EFSA report generated with the web tool).

B) Combined dataset (for males and females, BMR 20%):

BMDL: 49.3 mg/kg bw/d
BMDU: 83.2 mg/kg bw/d

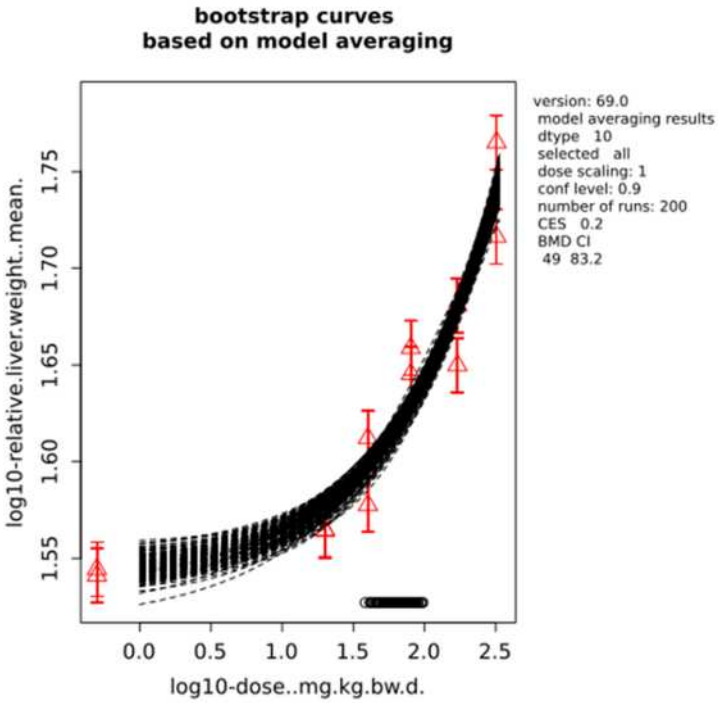


Figure 2-9 Graphical representation of the modelling results for the combined dataset (BMR 20%, taken from the EFSA report generated with the web tool).

Sperm motility

BMDL: 0.0038 mg/kg bw/d
BMDU: 34.9 mg/kg bw/d

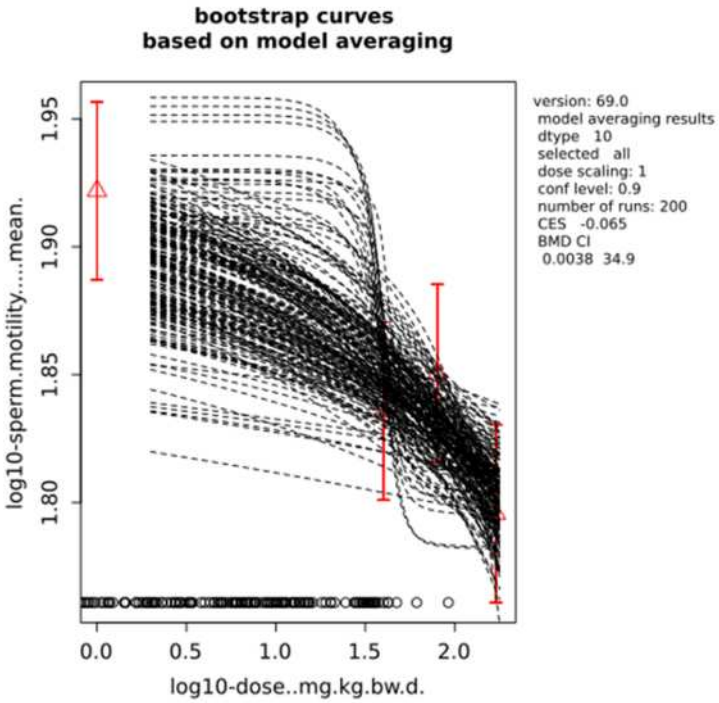


Figure 2-10 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool).

Discussion / Comparison with NOAEL:
For the endpoint “relative liver weight” modelling the combined dataset for males and females resulted in an “AIC warning”. This does not mean that the modelling results should not be used for the derivation of a POD, but warrants a detailed consideration of the results. For several reasons it was decided that the results could be used regardless of the “AIC warning”:

- Data clearly indicated a dose-dependency with an increasing trend starting in the first dose group for males and females
- Visual inspection of the graphs (see **Figure 2-8** and **Figure 2-9**) showed a good fit of the models.
- BMDU and BMDL have the same order of magnitude
- Modelling of the individual data for males and females separately led to comparable results.

The second modelling (“sperm motility) is not appropriate for the derivation of a POD in comparison with the results obtained for the endpoint “relative liver weight”. The uncertainty of the data is reflected in the large scattering of the bootstrap curves and the four orders of magnitude between BMDL and BMDU.

The following table shows a comparison of the NOAEL used by the MAK Commission (Hartwig 2020) (derived from the same study used here for benchmark modelling) for the derivation of an OEL and the BMDL calculated with the EFSA-tool.

Table 2-17 Comparison of NOAEL and BMDL

NOAEL (used by MAK Commission)	BMDL
20 mg/kg bw/d	49.3 mg/kg bw/d*

*results taken from the BMD modelling with a BMR of 20% increase in relative liver weight

2.2.3 N-octadecyl β-(3',5'-di-tert-butyl-4'-hydroxyphenyl) propionate (OBPP)

In the 14 days gavage toxicity study from Lake et al. (1980) induction of a number of parameters of hepatic microsomal xenobiotic metabolism and increase of relative liver weight were observed in male rats (see
Table **2-18**). For benchmark dose modelling the data on increase in relative liver weight in males were selected.

Table 2-18 Data on relative liver weight in male rats (according to Lake (1980)) used for benchmark dose modelling

dose (mg/kg bw/d)	Relative liver weight (mean in g /100 g bw)	Relative liver weight (SEM in g /100 g bw)	n # animals in group	sex
0	5	0.1	6	m
30	5.6	0.1	5	m
100	6	0.2	5	m
300	6.9	0.1	5	m
1000	7.9	0.3	5	m

m = male

Modelling parameters:

- Selection of BMR

  A) BMR 10% for changes in relative liver weight

  As mentioned already above for 1,1,2,2 tetrachloroethane, a BMR for changes in relative liver weight according to Dekkers et al. (2001) is recommended with 5% based on human data. In consistency to the approach selected for 1,1,2,2 tetrachloroethane above, the BMR was set at 2 SD (SEM was transferred to SD) of the control group:
  For male rats: 5.0 (100%) + 2 SD = 5.49 (109.8%)
  Therefore, a BMR of 10% was selected for modelling

B) BMR 20% for changes in relative liver weight

As outlined above in section 2.2.2 (1,1,2,2, tetrachloroethane) a second benchmark dose modelling was performed with a BMR of 20% for the endpoint “changes in relative liver weight”

- Model averaging

Results:
The complete reports generated with the EFSA web tool are included in the Annex.

A) BMR 10%

BMDL: 6.86 mg/kg bw/d
BMDU: 43.1 mg/kg bw/d

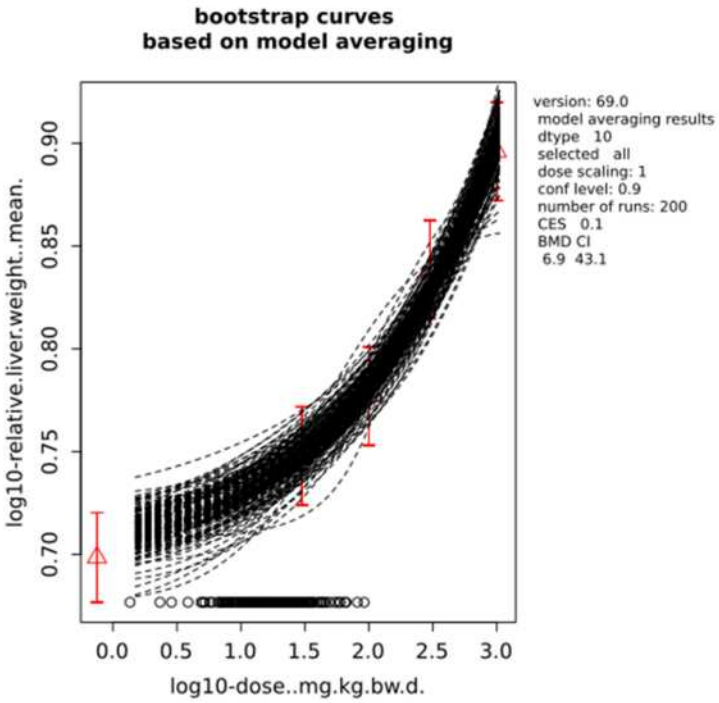


Figure 2-11 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool, BMR 10%).

B) BMR 20%

BMDL: 42.6 mg/kg bw/d
BMDU: 157 mg/kg bw/d

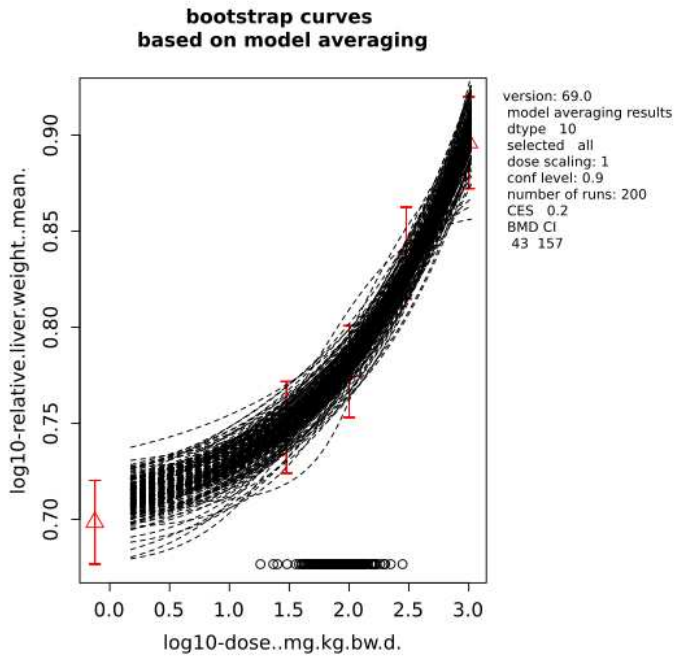


Figure 2-12 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool, BMR 20%).

Discussion / Comparison with NOAEL:
The following table shows a comparison of the NOAEL used by the MAK Commission (Hartwig and MAK Commission 2016) for the derivation of an OEL (coming from the same study used here for benchmark modelling) and the BMDL calculated with the EFSA-tool.

Table 2-19 Comparison of NOAEL and BMDL

NOAEL (used by MAK Commission)	BMDL
30 mg/kg bw/d	42.6 mg/kg bw/d

*results taken from the BMD modelling with a BMR of 20%

2.2.4 Tert-Butyl alcohol

In the MAK documentation from 2014 (Hartwig 2014) the increase of relative kidney weight in female rats observed in the chronic toxicity study from NTP (1995) was considered the most sensitive endpoint for the derivation of an OEL. The data are presented on the following
Table **2-20**. According to the NTP report, the data were obtained in the 15-month interim evaluation of the 2-year drinking water study.

Table 2-20 Data on relative kidney weight in female rats (according to NTP (1995)) used for benchmark dose modelling

dose (mg/kg bw/d)	Relative kidney weight (mean in mg organ weight / g bw)	Relative kidney weight (SEM in mg organ weight / g bw)	n # animals in group
0	3.49	0.08	10
180	3.99	0.07	10
330	4.21	0.08	10
650	4.95	0.17	10

Modelling parameters:

- Selection of BMR: In the BMD modelling reported in the MAK documentation for the substance (Hartwig 2014) a BMR of 1 SD was selected. To allow a comparison of the results, the same BMR was selected for the modelling presented here. Since the BMR in the EFSA web tool can only be selected in form of percentages, the control value + 1 SD (SD calculated from SEM) was compared to the control value and the change in percent (+ 7%) was determined. Therefore, a BMR of 7% was selected. Dekkers et al. do not indicate a BMR for changes in relative kidney weight.

- Model averaging

Results:
The complete reports generated with the EFSA web tool are included in the Annex.

BMDL: 47.7 mg/kg bw/d
BMDU: 181 mg/kg bw/d

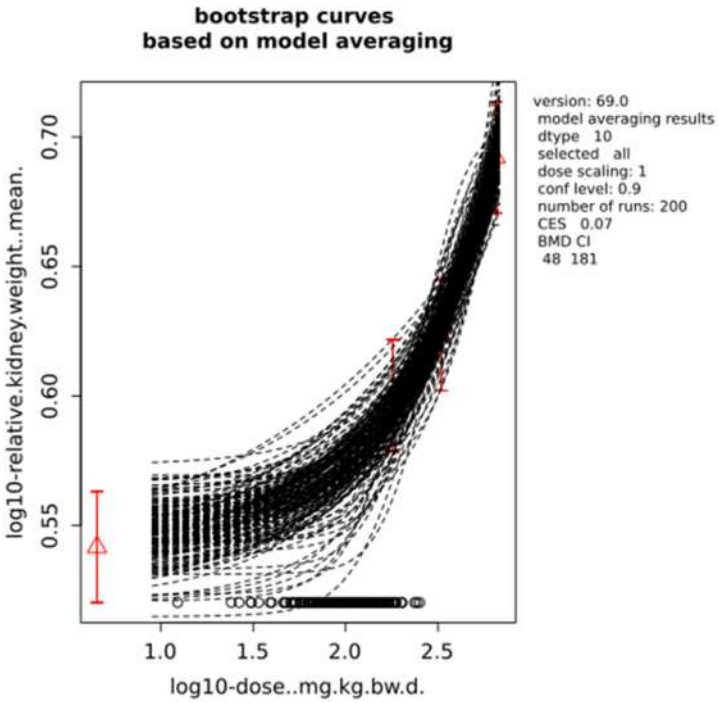


Figure 2-13 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool).

Discussion / Comparison with other BMDLs:
The following table shows a comparison of the BMDL used by the MAK Commission for the derivation of an OEL (coming from the same study used here for benchmark modelling) and the BMDL calculated with the EFSA-tool.

Table 2-21 Comparison of NOAEL and BMDL

BMDL (used by MAK Commission)*	BMDL
68 mg/kg bw/d	47.7 mg/kg bw/d

*In the MAK documentation (Hartwig 2014) a BMD-modelling was performed with BMDS (v1.4.1) and the "polynomial" model was selected resulting in a BMD of 92 mg/kg bw/d and a BMDL of 68 mg/kg bw/d.

2.2.5 Benzene

In an epidemiological study blood from exposed workers and controls was analysed for the number of white blood cells (Zhang et al. 2016). The data are presented in the following table.

Table 2-22 Data on white blood cell count in exposed workers and controls (according to Zhang et al. (2016))

dose [ppm-year]	WBC count (mean x10^9/L)	SD	n (number of workers)
0	6.48	1.42	94
3.55	6.14	1.6	65
6.51	6.14	1.33	65
10.72	5.76**	1.57	65
20.02	6.04*	1.87	65
40.71	5.7**	1.6	65

* P < 0.05; ** P < 0.01

Modelling parameters:

- According to the procedure selected by Zhang et al. (2016) the continuous data presented above were transferred to quantal data using a reduced white blood cell (WBC) count less or equal to the 5th percentile of the control distribution. This corresponds to a value of 4.3 x 10^9 cells/L. The adverse effect (leukocytopenia) manifests with clinical effects around 4.00 x 10^9 WBC/L (Medizinische Fachredaktion Pschyrembel 2018). Using “simple computation” the “abnormality” N was calculated by Zhang et al (see the following table).

Table 2-23 Data on white blood cell count in exposed workers and controls (according to Zhang (2016)) with a column (“abnormality”) presenting the data transferred to a quantal presentation

dose [ppm-year]	WBC count (mean x10^9/L)	SD	n	Abnormality N (%)
0	6.48	1.42	94	4 (4.3)
3.55	6.14	1.6	65	7 (10.8)
6.51	6.14	1.33	65	6 (9.2)
10.72	5.76	1.57	65	14 (21.5)
20.02	6.04	1.87	65	16 (24.6)
40.71	5.7	1.6	65	16 (24.6)

- Zhang et al. used a BMR of 5 and 10% to model the quantal dataset (dose, n, and N). A BMR of 10% was also selected for BMD-modelling in this report.

- Model averaging

Results:
The complete reports generated with the EFSA web tool are included in the Annex.

BMDL: 1.3 ppm-years
BMDU: 19 ppm-years

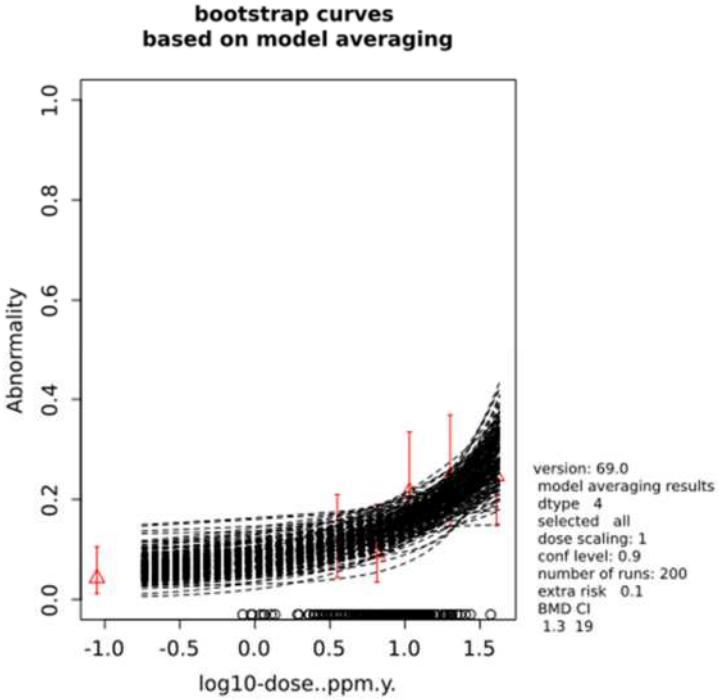


Figure 2-14 Graphical representation of the modelling results (taken from the EFSA report generated with the web tool).

Discussion / Comparison with other BMDLs:
The following table shows a comparison of the NOAEL derived by the authors of this document based on the significance of effects as indicated by Zhang et al. and the BMDL calculated with the EFSA-tool.
Setting a BMR for this dataset is a special case: On an individual level it is well established that a decrease of the number of white blood cells to approx. 4 x 10^9 cells/L constitutes the borderline for clinical concerns. In the human population in general and also in the study group of Zhang et al. a large variability in the individual counts of white blood cell can be observed (observable from the large SD in **Table 2-22**, which means that the distance to the critical cell count is highly different from individual to individual. Setting the BMR for the group mean would result in a dose, at which 50% of the population would have WBC counts indicating clinical concerns. Therefore, the transformation to quantal data according to Zhang et al. (2016) using a reduced WBC count less or equal to the 5th percentile of the control distribution was adopted and a BMR associated with an incidence level of 10% was used.

Table 2-24 Comparison of NOAEL and BMDL

NOAEL (derived by authors of this report)	BMDL
6.51 ppm-years	1.3 ppm-years

References

Cho, W.-S., Han, B.S., Nam, K.T., Park, K., Choi, M., Kim, S.H., Jeong, J., Jang, D.D. (2008). Carcinogenicity study of 3-monochloropropane-1,2-diol in Sprague-Dawley rats. Food and Chemical Toxicology 46, 3172-3177.

Dekkers, S., de Heer, C., Rennen, M.A.J. (2001). Critical effect sizes in toxicological risk assessment: a comprehensive and critical evaluation. Environmental Toxicology and Pharmacology 10, 33-52.

EFSA Scientific Committee, Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, K.H., More, S., Mortensen, A., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Silano, V., Solecki, R., Turck, D., Aerts, M., Bodin, L., Davis, A., Edler, L., Gundert-Remy, U., Sand, S., Slob, W., Bottex, B., Abrahantes, J.C., Marques, D.C., Kass, G., Schlatter, J.R. (2017). Update: use of the benchmark dose approach in risk assessment. The EFSA Journal 15(1):4658www, 41pp.

Greim, H. (2008). Gesundheitsschädliche Arbeitsstoffe, Toxikologisch-arbeitsmedizinische Begründungen von MAK-Werten, Loseblattsammlung, 45. Lfg, DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag Weinheim.

Greim, H., MAK Kommission. (2008). Nitrilotriessigsäure und ihre Natriumsalze. Gesundheitsschädliche Arbeitsstoffe, Toxikologisch-arbeitsmedizinische Begründungen von MAK-Werten, Loseblattsammlung, 45. Lfg, DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag Weinheim. https://onlinelibrary.wiley.com/doi/pdf/10.1002/3527600418.mb13913verd0045, p. 72pp.

Haber, L.T., Dourson, M.L., Allen, B.C., Hertzberg, R.C., Parker, A., Vincent, M.J., Maier, A., Boobis, A.R. (2018). Benchmark dose (BMD) modeling: current practice, issues, and challenges. Critical Reviews in Toxicology 48, 387-415.

Hartwig, A. (2014). Gesundheitsschädliche Arbeitsstoffe, Toxikologisch-arbeitsmedizinische Begründungen von MAK-Werten, Loseblattsammlung, 56. Lfg, DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag Weinheim
Hartwig, A. (2020). 1,1,2,2-Tetrachlorethan, Nachtrag 2020, Entwurfsfassung. Gesundheitsschädliche Arbeitsstoffe, Toxikologisch-arbeitsmedizinische Begründungen von MAK-Werten, Loseblattsammlung. DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag Weinheim.

Hartwig, A., MAK Commission. (2016). 3,5-Di-tert-butyl-4-hydroxyphenylpropionsäureoctadecylester [MAK Value Documentation in German language]. The MAK Collection for Occupational Health and Safety 2016, Vol. 1, No. 2, DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag GmbH & Co. KGaA, pp. 887-916.

Hartwig, A., MAK Commission. (2018). Benzoic acid and alkali benzoates. The MAK Collection for Occupational Health and Safety 2017, Vol. 3, No. 4, DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag GmbH & Co. KGaA.

Kommineni, C., Groth, D.H., Frockt, I.J., Voelker, R.W., Stanovick, R.P. (1979). Determination of the tumorigenic potential of methylene-bis-orthochloroaniline. Journal of Environmental Pathology and Toxicology 2, 149-171.

Lake, B.G., Gangolli, S.D., Schmid, K., Schweizer, W., Stäubli, W., Waechter, F. (1980). The induction of rat hepatic microsomal xenobiotic metabolism by n-octadecyl β-(3′,5′-di-tert-butyl-4′-hydroxyphenyl)-propionate. Food and Cosmetics Toxicology 18, 47-54.

Medizinische Fachredaktion Pschyrembel. (2018). Pschyrembel Online, Walter de Gruyter GmbH. https://www.pschyrembel.de/.

NTP, National Toxicology Program. (1989). Toxicology and Carcinogenesis Studies of Nalidixic Acid in F344/N Rats and B6C3F1 Mice (Feed Studies). TR 368, U.S. Department of Health and Human Services; Public Health Service.

NTP, National Toxicology Program. (1995). Toxicology and Carcinogenesis Studies of t-Butyl Alcohol in F344/N Rats and B6C3F1 Mice (Drinking Water Studies). TR 436, U.S. Department of Health and Human Services; Public Health Service.

NTP, National Toxicology Program. (2002). Toxicology and Carcinogenesis Studies of Vanadium Pentoxide in F344/N Rats and B6C3F1 Mice (Inhalation Studies). TR 507, U.S. Department of Health and Human Services; Public Health Service.

NTP, National Toxicology Program. (2004). Technical Report on the Toxicity Studies of 1,1,2,2-Tetrachloroethane Administered in Microcapsules in Feed to F344/N Rats and B6C3F1 Mice. Toxicity Report Series 49, U.S. Department of Health and Human Services; Public Health Service.
https://ntp.niehs.nih.gov/ntp/htdocs/st_rpts/tox049.pdf.

RAC, Committee for Risk Assessment. (2017). Opinion on 4,4'-methylene-bis-[2-chloroaniline] (MOCA). ECHA/RAC/A77-O-0000001412-86-147/F. Adopted 29 May 2017, European Chemicals Agency (ECHA), Helsinki, Finland.
https://echa.europa.eu/documents/10162/13641/opinion_moca_en.pdf/35756093-0eb9-e468-2ba2-786ca73c5aaa.

WHO, World Health Organization. (2009). Environmental Health Criteria 240, Principles and Methods for the Risk Assessment of Chemicals in Food, IPCS International Programme on Chemical Safety; World Health Organization, Geneva, Switzerland.

WHO, World Health Organization. (2020). Environmental Health Criteria 240, Principles and Methods for the Risk Assessment of Chemicals in Food. Chapter 5: Dose–response assessment and derivation of health-based guidance values, IPCS

International Programme on Chemical Safety; World Health Organization, Geneva. https://www.who.int/docs/default-source/food-safety/publications/chapter5-dose-response.pdf?sfvrsn=32edc2c6_5.

Zhang, G.H., Ji, B.Q., Li, Y., Zheng, G.Q., Ye, L.L., Hao, Y.H., Ren, J.C., Zhou, L.F., Xu, X.W., Zhu, Y., Xia, Z.L. (2016). Benchmark doses based on abnormality of WBC or micronucleus frequency in benzene-exposed Chinese workers. Journal of Occupational and Environmental Medicine 58, e39-44.

List of Tables

List of Figures

Annex: BMD modelling protocols

BMD modelling was performed with the EFSA web tool available at https://r4eu.efsa.europa.eu/.

Benchmark Dose Modeling: Report for 3-Monochloropropane-1,2-diol (3-MCPD)

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: effect.

Data used for analysis:

dose	effect	n
0.00	1	50
1.97	11	50
8.27	21	50
29.50	36	50

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is an extra risk of 10% compared to the controls.

When the specified BMR deviates from the default value, the rationale behind the choice made should be described.

The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 67.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

Default set of fitted models:

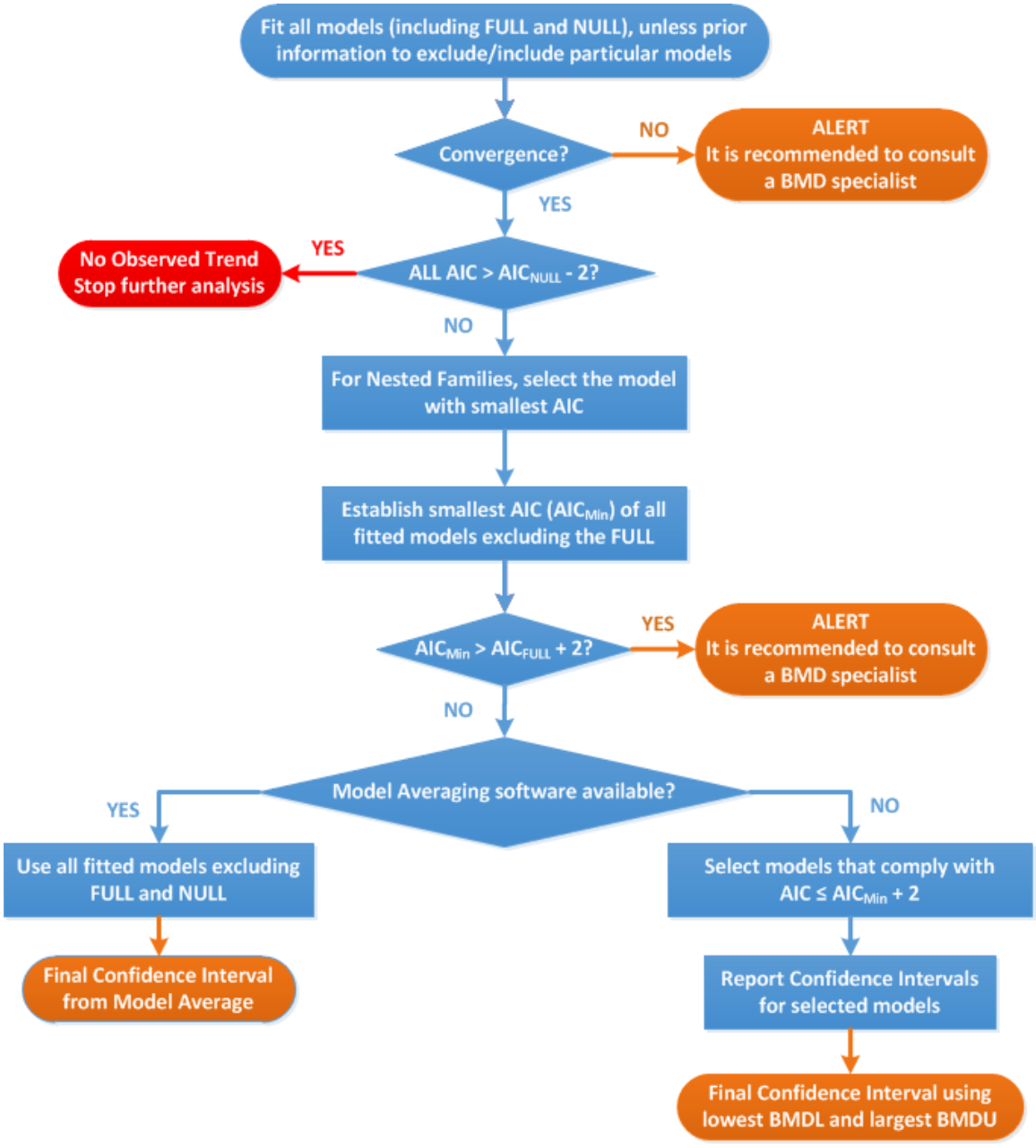
Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean
Logistic	2	$y = \frac{1}{1 + \exp(-a - bx)}$
Probit	2	$y = pnorm((x - a) \cdot b)$

Log-logistic	3	$y = a + \frac{1-a}{1+\exp\left(c \cdot \log\left(\frac{b}{x}\right)\right)}$
Log-probit	3	$y = a + (1-a) \cdot pnorm\left(c \cdot \log\left(\frac{x}{b}\right)\right)$
Weibull	3	$y = a + (1-a)\left(1 - \exp\left(-\left(\frac{x}{b}\right)^{c}\right)\right)$
Gamma	3	$y = pgamma(bx; c)$
Two-stage	3	$y = a + (1 - a)\left(1 - \exp\left(-\frac{x}{b} - c\left(\frac{x}{b}\right)^{2}\right)\right)$
Exp model 3	3	$y = a \cdot \exp(bx^{d})$
Exp model 5	4	$y = a \cdot (c - (c-1)\exp(-bx^{d}))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^{d}}{b^{d} + x^{d}}\right)$
Hill model 5	4	$y = a \cdot \left(1 + (c-1)\frac{x^{d}}{b^{d} + x^{d}}\right)$

For the Exp and Hill family, we fit models with 3 and 4 parameters as listed in the table. The 3-parameter model is selected if the difference in AIC is smaller than 5, otherwise the 4-parameter model is selected.

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: effect

Fitted Models

model	No.par	loglik	AIC	accepted	BMDL	BMDU	BMD	conv
null	1	-128.86	259.72		NA	NA	NA	NA
full	4	-94.91	197.82		NA	NA	NA	NA
two.stage	3	-97.49	200.98	no	NA	NA	2.140	yes
log.logist	3	-95.07	196.14	yes	0.224	1.88	0.831	yes
Weibull	3	-94.94	195.88	yes	0.135	1.64	0.631	yes
log.prob	3	-95.10	196.20	yes	0.276	1.93	0.917	yes
gamma	3	-94.92	195.84	yes	0.074	1.60	0.526	yes
logistic	2	-102.50	209.00	no	NA	NA	5.620	yes
probit	2	-102.20	208.40	no	NA	NA	5.360	yes
LVM: Expon. m3-	3	-94.92	195.84	yes	0.171	1.39	0.484	yes
LVM: Hill m3-	3	-94.94	195.88	yes	0.131	1.53	0.575	yes

Estimated Model Parameters

two.stage

estimate for a- : 0.05307

estimate for BMD- : 2.138

estimate for c : 1e-06

log.logist

estimate for a- : 0.02061

estimate for BMD- : 0.831

estimate for c : 0.8499

Weibull

estimate for a- : 0.02025

estimate for BMD- : 0.6308

estimate for c : 0.6401

log.prob

estimate for a- : 0.02051

estimate for BMD- : 0.9169

estimate for c : 0.5174

gamma

estimate for a- : 0.02008

estimate for BMD- : 0.5264

estimate for cc : 0.5359

logistic

estimate for a- : -1.71

estimate for BMD- : 5.623

probit

estimate for a- : -1.051

estimate for BMD- : 5.356

EXP

estimate for a- : 1.669

estimate for CED- : 0.4839

estimate for d- : 0.2702

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

HILL

estimate for a- : 1.666

estimate for CED- : 0.5747

estimate for d- : 0.3416

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

Weights for Model Averaging

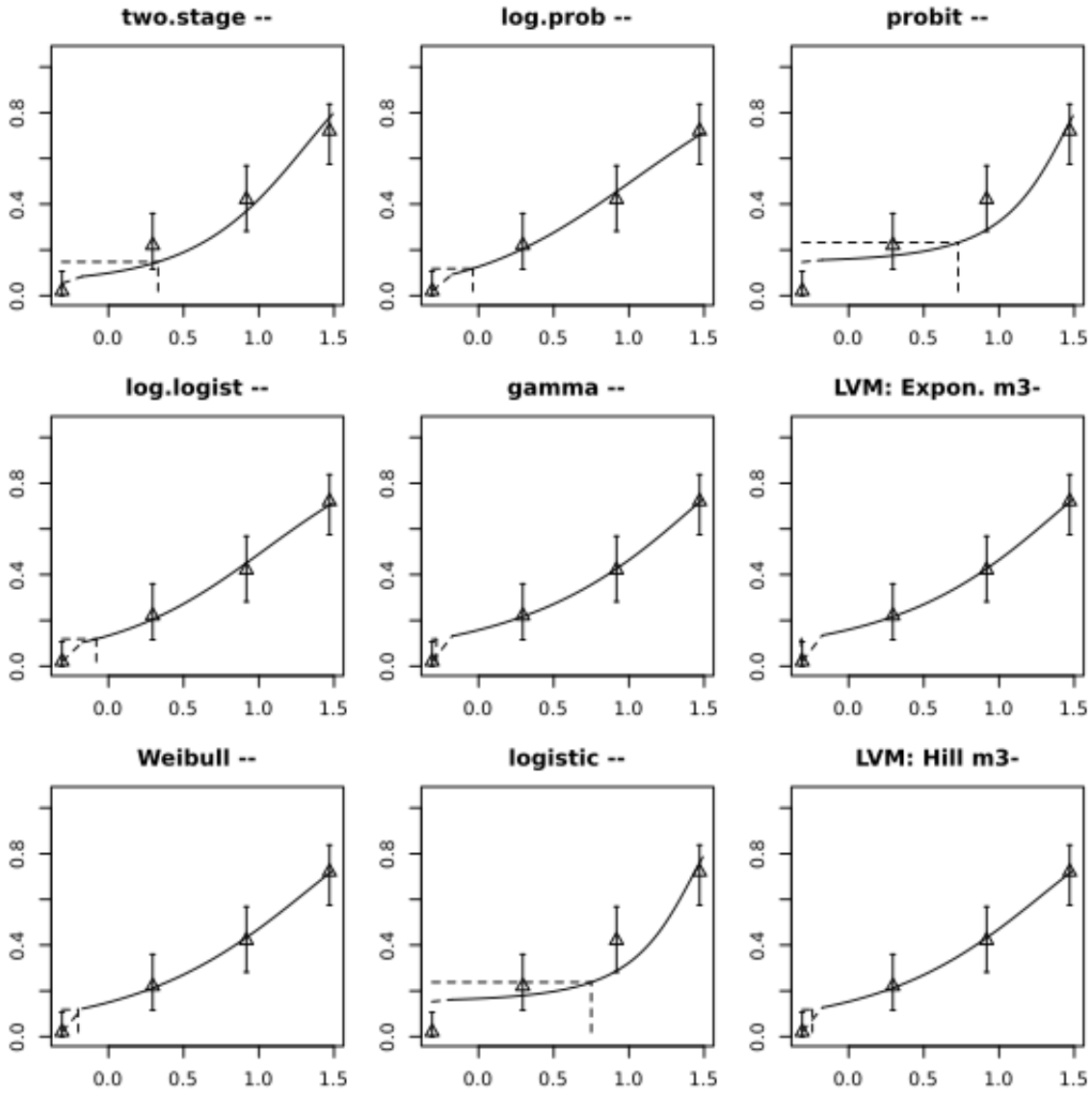
two.stage	log.logist	Weibull	log.prob	gamma	logistic	probit	EXP	HILL
0.01	0.15	0.17	0.15	0.17	0	0	0.17	0.17

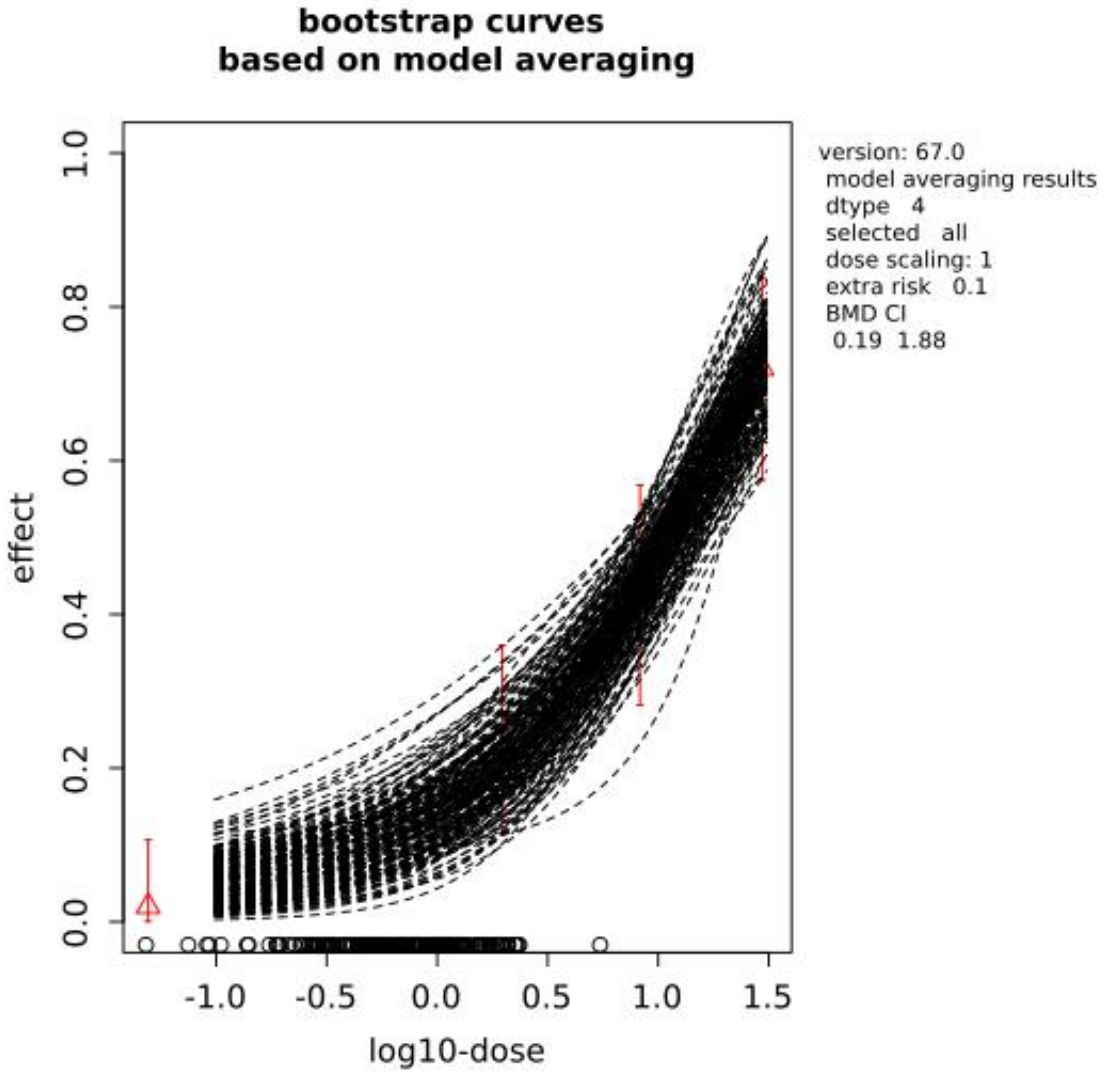
Final BMD Values

subgroup	BMDL	BMDU
	0.19	1.88

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization





Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for Vanadium pentaoxide

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: effect..Lunge..

Data used for analysis:

	dose	effect..Lunge.	n
1	0.00	5	50
5	0.00	10	49
2	0.28	8	49
6	0.28	10	49
3	0.56	24	48
7	0.56	14	50
4	1.12	42	50
8	1.12	40	50

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is an extra risk of 10% compared to the controls.

When the specified BMR deviates from the default value, the rationale behind the choice made should be described.

The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

Default set of fitted models:

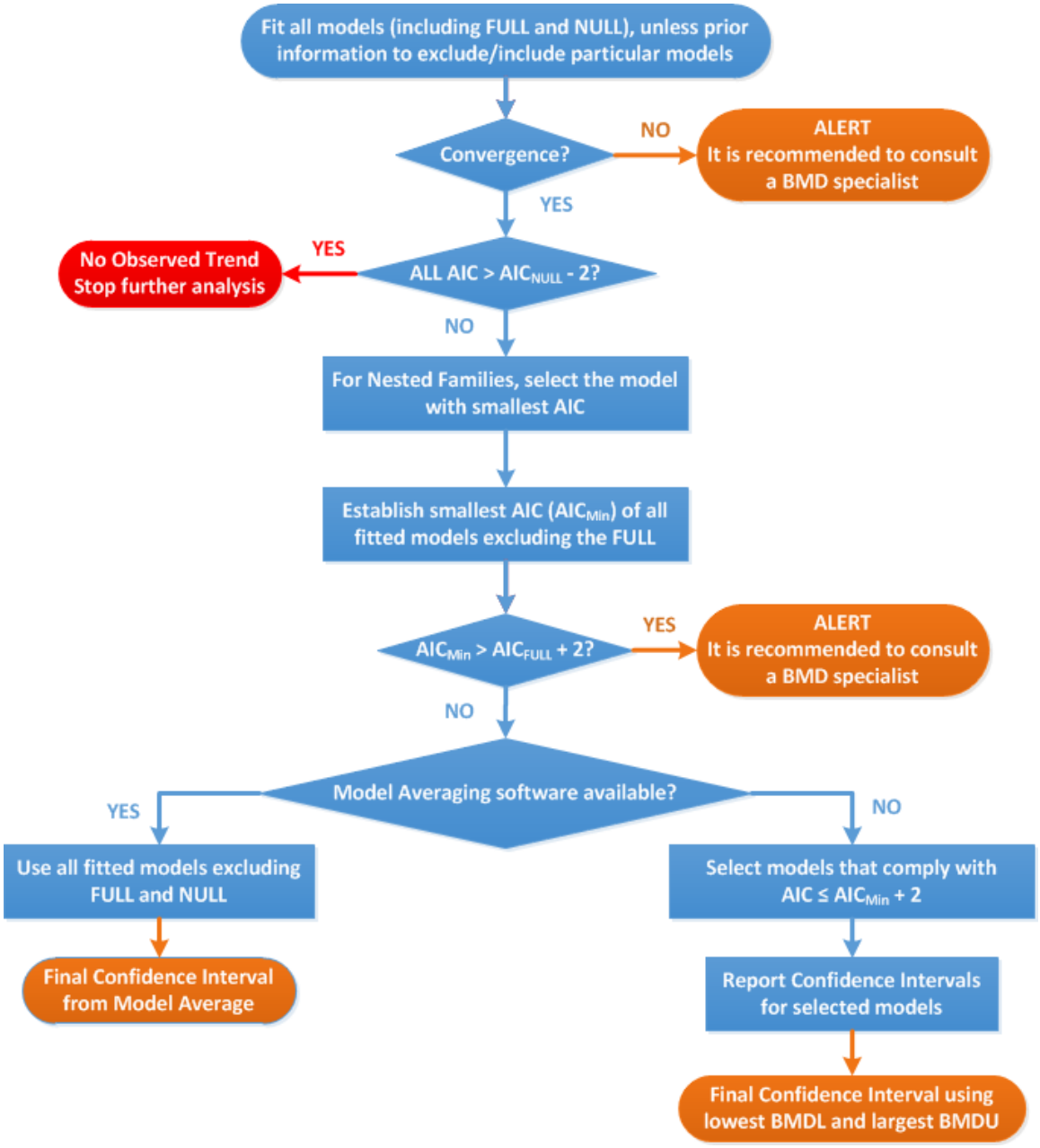
Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean

Logistic	2	$y = \frac{1}{1 + \exp(-a - bx)}$
Probit	2	$y = pnorm((x - a) \cdot b)$
Log-logistic	3	$y = a + \frac{1 - a}{1 + \exp\left(c \cdot \log\left(\frac{b}{x}\right)\right)}$
Log-probit	3	$y = a + (1 - a) \cdot pnorm\left(c \cdot \log\left(\frac{x}{b}\right)\right)$
Weibull	3	$y = a + (1 - a)\left(1 - \exp\left(-\left(\frac{x}{b}\right)^{c}\right)\right)$
Gamma	3	$y = pgamma(bx; c)$
Two-stage	3	$y = a + (1 - a)\left(1 - \exp\left(-\frac{x}{b} - c\left(\frac{x}{b}\right)^{2}\right)\right)$
Exp model 3	3	$y = a \cdot \exp(bx^{d})$
Exp model 5	4	$y = a \cdot (c - (c - 1)\exp(-bx^{d}))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^{d}}{b^{d} + x^{d}}\right)$
Hill model 5	4	$y = a \cdot \left(1 + (c - 1)\frac{x^{d}}{b^{d} + x^{d}}\right)$

For the Exp and Hill family, we fit models with 3 and 4 parameters as listed in the table. The 3-parameter model is selected if the difference in AIC is smaller than 5, otherwise the 4-parameter model is selected.

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: effect..Lunge.

Fitted Models

model	No.par	loglik	AIC	accepted	BMDL	BMDU	BMD	conv
null	1	-263.68	529.36		NA	NA	NA	NA
full	4	-201.42	410.84		NA	NA	NA	NA
two.stage	3	-201.90	409.80	yes	0.269	0.330	0.297	yes
log.logist	3	-201.42	408.84	yes	0.288	0.485	0.383	yes
Weibull	3	-201.53	409.06	yes	0.250	0.469	0.348	yes
log.prob	3	-201.45	408.90	yes	0.297	0.495	0.394	yes
gamma	3	-201.43	408.86	yes	0.271	0.485	0.372	yes
logistic	2	-203.03	410.06	yes	0.199	0.258	0.226	yes
probit	2	-203.45	410.90	no	NA	NA	0.210	yes
LVM: Expon. m3-	3	-201.78	409.56	yes	0.216	0.458	0.320	yes
LVM: Hill m3-	3	-201.67	409.34	yes	0.231	0.462	0.331	yes

Estimated Model Parameters

two.stage

estimate for a- : 0.1323

estimate for BMD- : 0.297

estimate for c : 1e+12

log.logist

estimate for a- : 0.1513

estimate for BMD- : 0.3826

estimate for c : 3.268

Weibull

estimate for a- : 0.145

estimate for BMD- : 0.348

estimate for c : 2.312

log.prob

estimate for a- : 0.156

estimate for BMD- : 0.3944

estimate for c : 1.985

gamma

estimate for a- : 0.1499

estimate for BMD- : 0.3721

estimate for c : 4.206

logistic

estimate for a- : -2.101

estimate for BMD- : 0.2263

probit

estimate for a- : -1.244

estimate for BMD- : 0.2103

EXP

estimate for a- : 1.308

estimate for BMD- : 0.3201

estimate for d- : 1.449

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

HILL

estimate for a- : 1.306

estimate for BMD- : 0.3313

estimate for d- : 1.671

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

Weights for Model Averaging

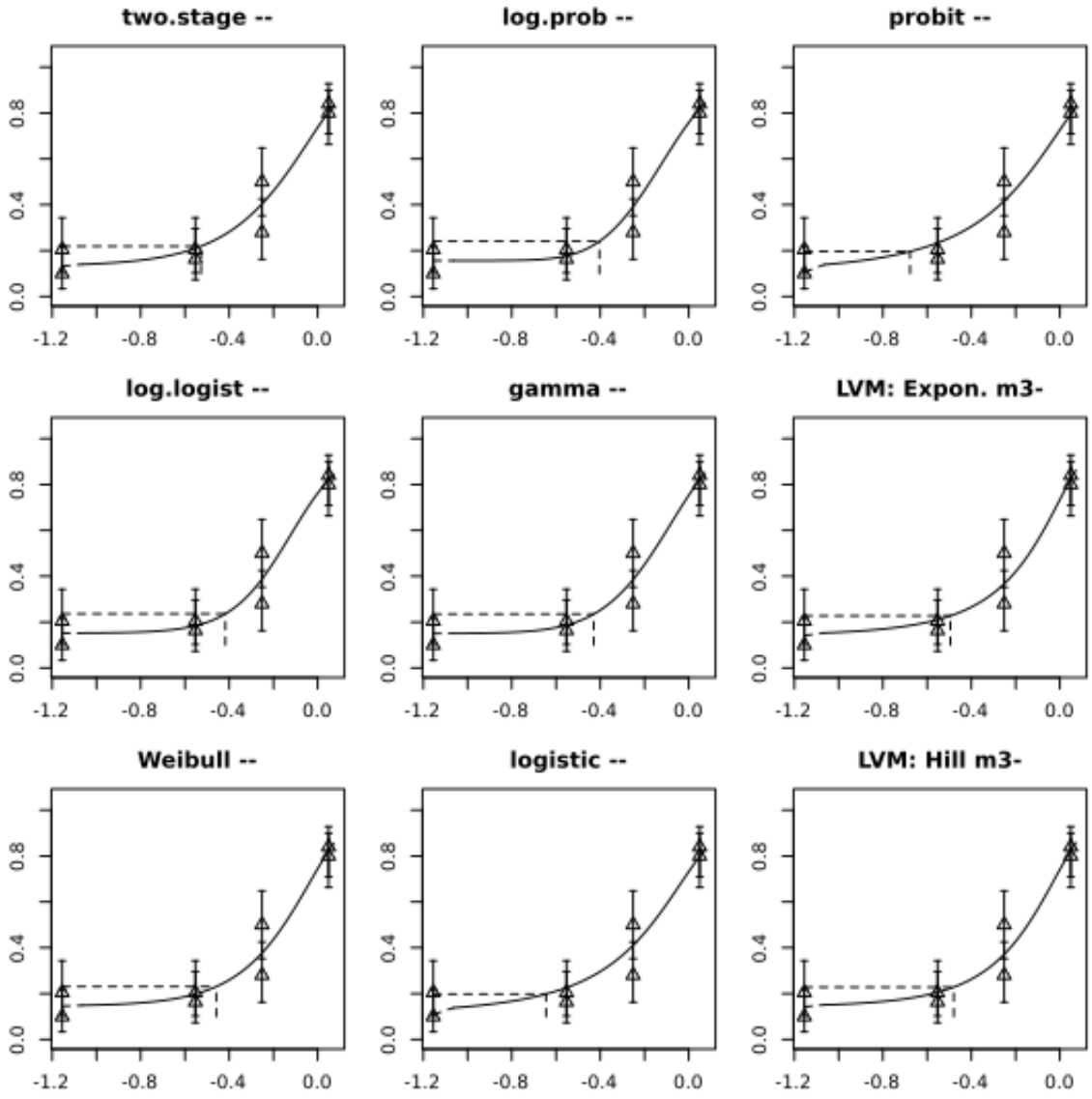
two.stage	log.logist	Weibull	log.prob	gamma	logistic	probit	EXP	HILL
0.09	0.15	0.13	0.14	0.14	0.08	0.05	0.1	0.11

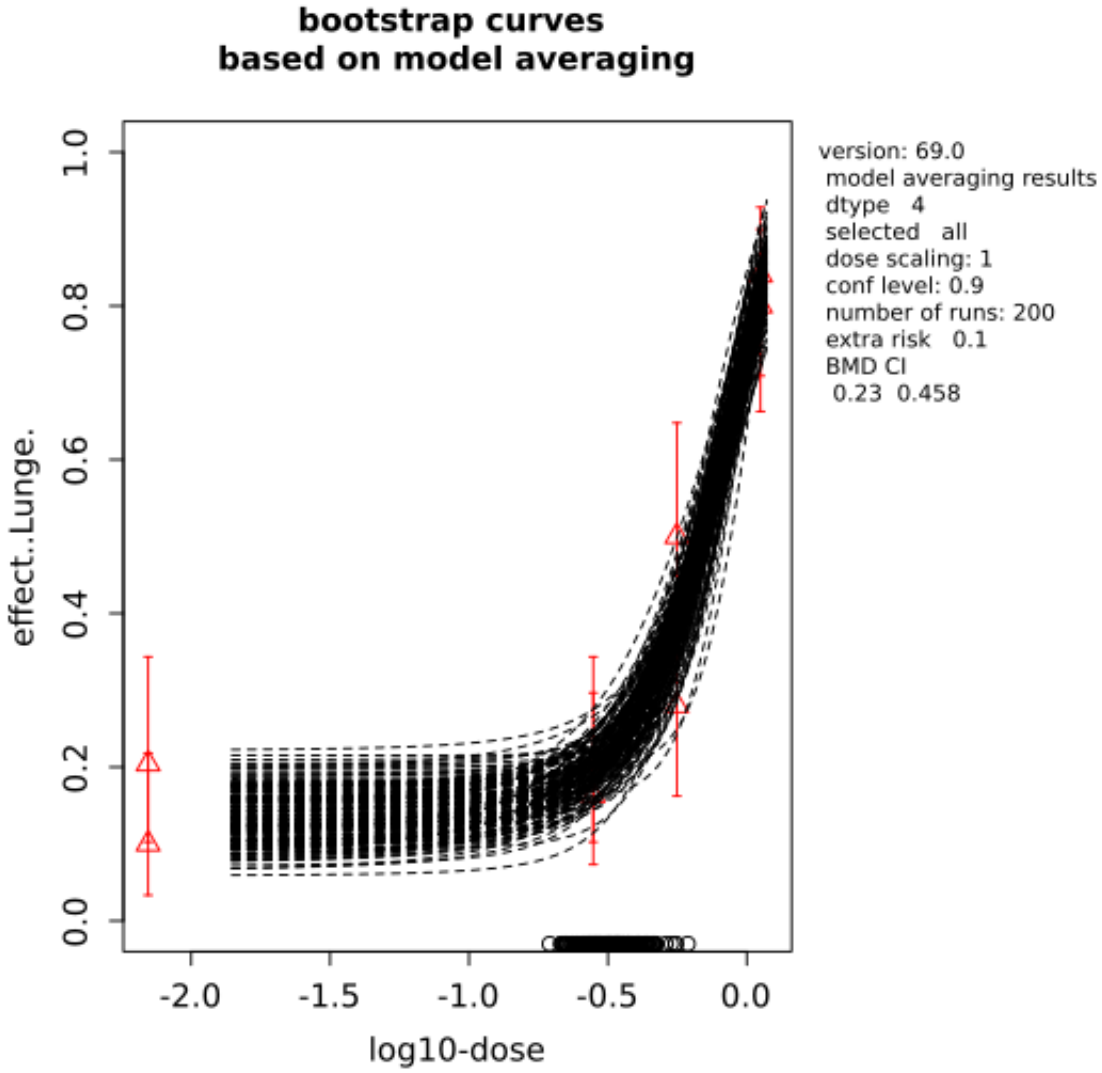
Final BMD Values

subgroup	BMDL	BMDU
all	0.23	0.46

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization





Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for 4,4’-Methylene-bis-[2-chloroaniline] (MOCA)

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: lung.tumours.

Data used for analysis:

corrected.dose..mg.kg.bw.d.	lung.tumours	n
0.0	1	100
9.4	23	100
18.8	28	75
37.5	35	50

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is an extra risk of 10% compared to the controls.

When the specified BMR deviates from the default value, the rationale behind the choice made should be described.

The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 67.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

Default set of fitted models:

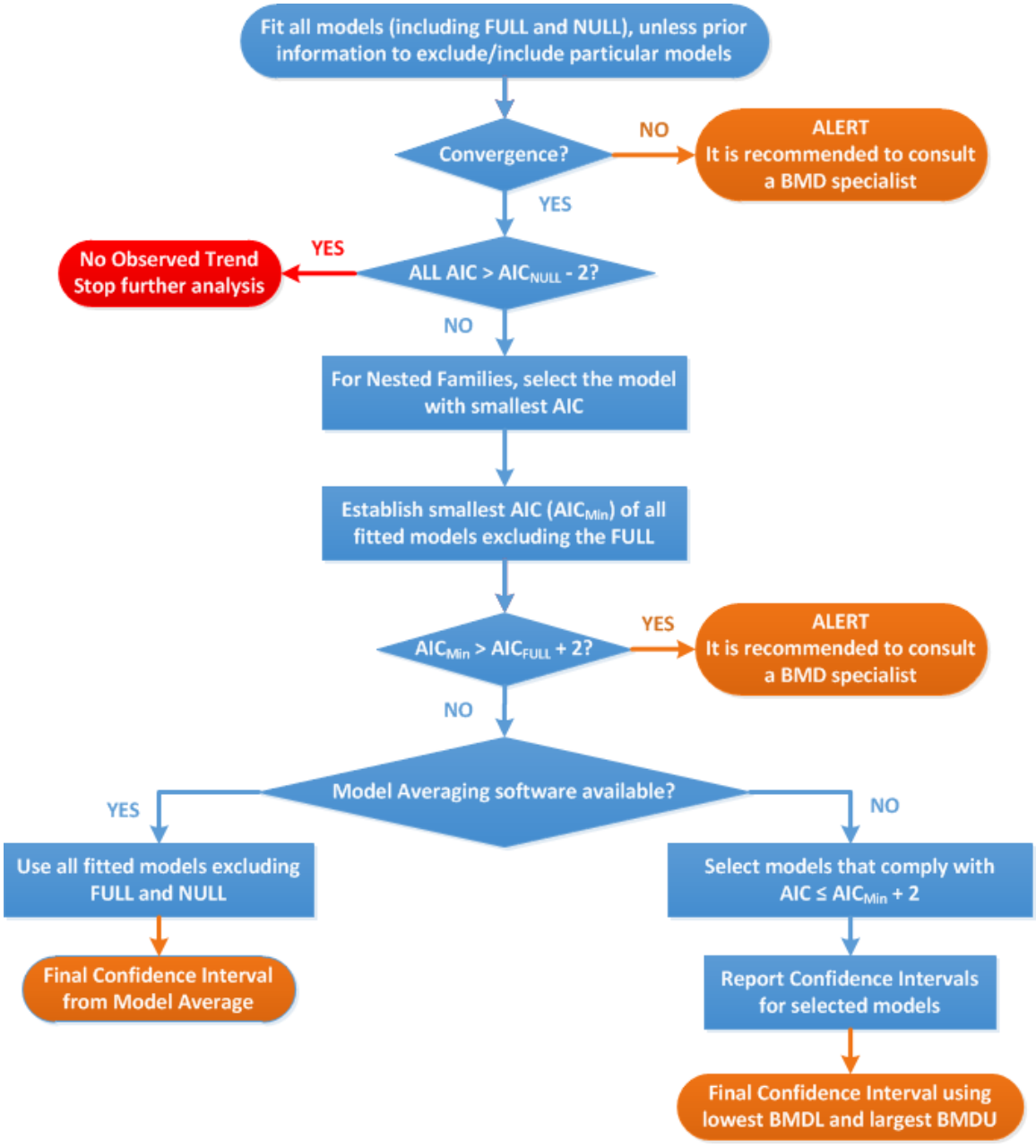
Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean
Logistic	2	$y = \frac{1}{1 + \exp(-a - bx)}$
Probit	2	$y = pnorm((x - a) \cdot b)$

Log-logistic	3	$y = a + \frac{1-a}{1+\exp\left(c \cdot \log\left(\frac{b}{x}\right)\right)}$
Log-probit	3	$y = a + (1-a) \cdot pnorm\left(c \cdot \log\left(\frac{x}{b}\right)\right)$
Weibull	3	$y = a + (1-a)\left(1 - \exp\left(-\left(\frac{x}{b}\right)^{c}\right)\right)$
Gamma	3	$y = pgamma(bx; c)$
Two-stage	3	$y = a + (1-a)\left(1 - \exp\left(-\frac{x}{b} - c\left(\frac{x}{b}\right)^{2}\right)\right)$
Exp model 3	3	$y = a \cdot \exp(bx^{d})$
Exp model 5	4	$y = a \cdot (c - (c-1)\exp(-bx^{d}))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^{d}}{b^{d} + x^{d}}\right)$
Hill model 5	4	$y = a \cdot \left(1 + (c-1)\frac{x^{d}}{b^{d} + x^{d}}\right)$

For the Exp and Hill family, we fit models with 3 and 4 parameters as listed in the table. The 3-parameter model is selected if the difference in AIC is smaller than 5, otherwise the 4-parameter model is selected.

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: lung.tumours

Fitted Models

model	No.par	loglik	AIC	accepted	BMDL	BMDU	BMD	conv
null	1	-188.81	379.62		NA	NA	NA	NA
full	4	-139.62	287.24		NA	NA	NA	NA
two.stage	3	-139.83	285.66	yes	3.26	6.59	4.46	yes
log.logist	3	-140.21	286.42	yes	3.01	7.53	5.33	yes
Weibull	3	-139.91	285.82	yes	2.49	7.01	4.72	yes
log.prob	3	-140.27	286.54	yes	3.36	7.74	5.66	yes
gamma	3	-139.94	285.88	yes	2.28	7.19	4.76	yes
logistic	2	-146.23	296.46	no	NA	NA	9.38	yes
probit	2	-145.21	294.42	no	NA	NA	8.80	yes
LVM: Expon. m3-	3	-139.84	285.68	yes	2.13	6.55	4.27	yes
LVM: Hill m3-	3	-139.93	285.86	yes	2.40	6.87	4.61	yes

Estimated Model Parameters

two.stage

estimate for a- : 0.01027

estimate for BMD- : 4.464

estimate for c : 0.4249

log.logist

estimate for a- : 0.01029

estimate for BMD- : 5.329

estimate for c : 1.469

Weibull

estimate for a- : 0.01021

estimate for BMD- : 4.719

estimate for c : 1.145

log.prob

estimate for a- : 0.01023

estimate for BMD- : 5.66

estimate for c : 0.8951

gamma

estimate for a- : 0.01017

estimate for BMD- : 4.763

estimate for cc : 1.208

logistic

estimate for a- : -2.546

estimate for BMD- : 9.384

probit

estimate for a- : -1.538

estimate for BMD- : 8.803

EXP

estimate for a- : 1.779

estimate for CED- : 4.268

estimate for d- : 0.4377

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

HILL

estimate for a- : 1.777

estimate for CED- : 4.608

estimate for d- : 0.5613

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

Weights for Model Averaging

two.stage	log.logist	Weibull	log.prob	gamma	logistic	probit	EXP	HILL
0.16	0.11	0.15	0.11	0.15	0	0	0.16	0.15

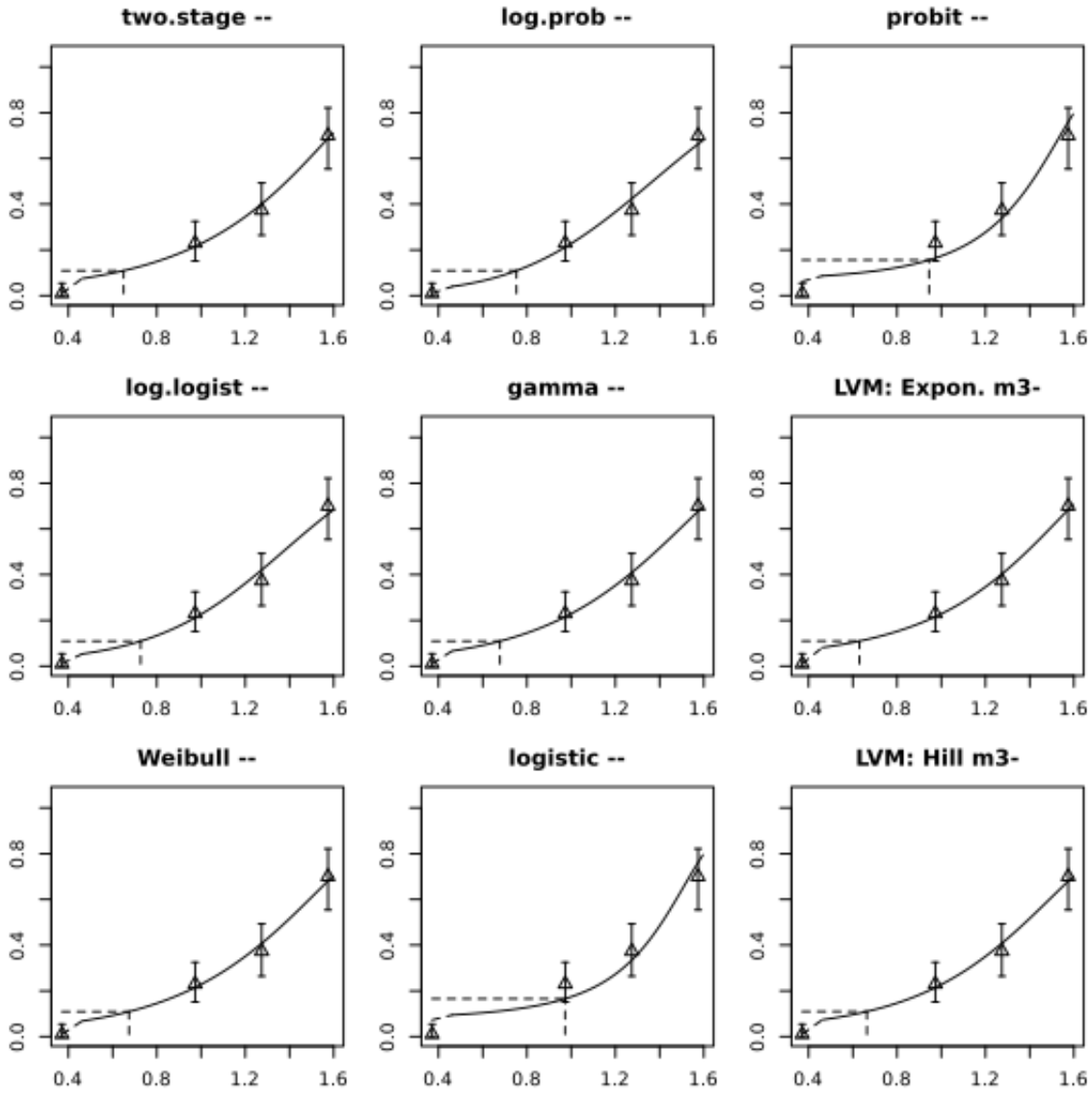
Final BMD Values

subgroup	BMDL	BMDU
	2.91	7.19

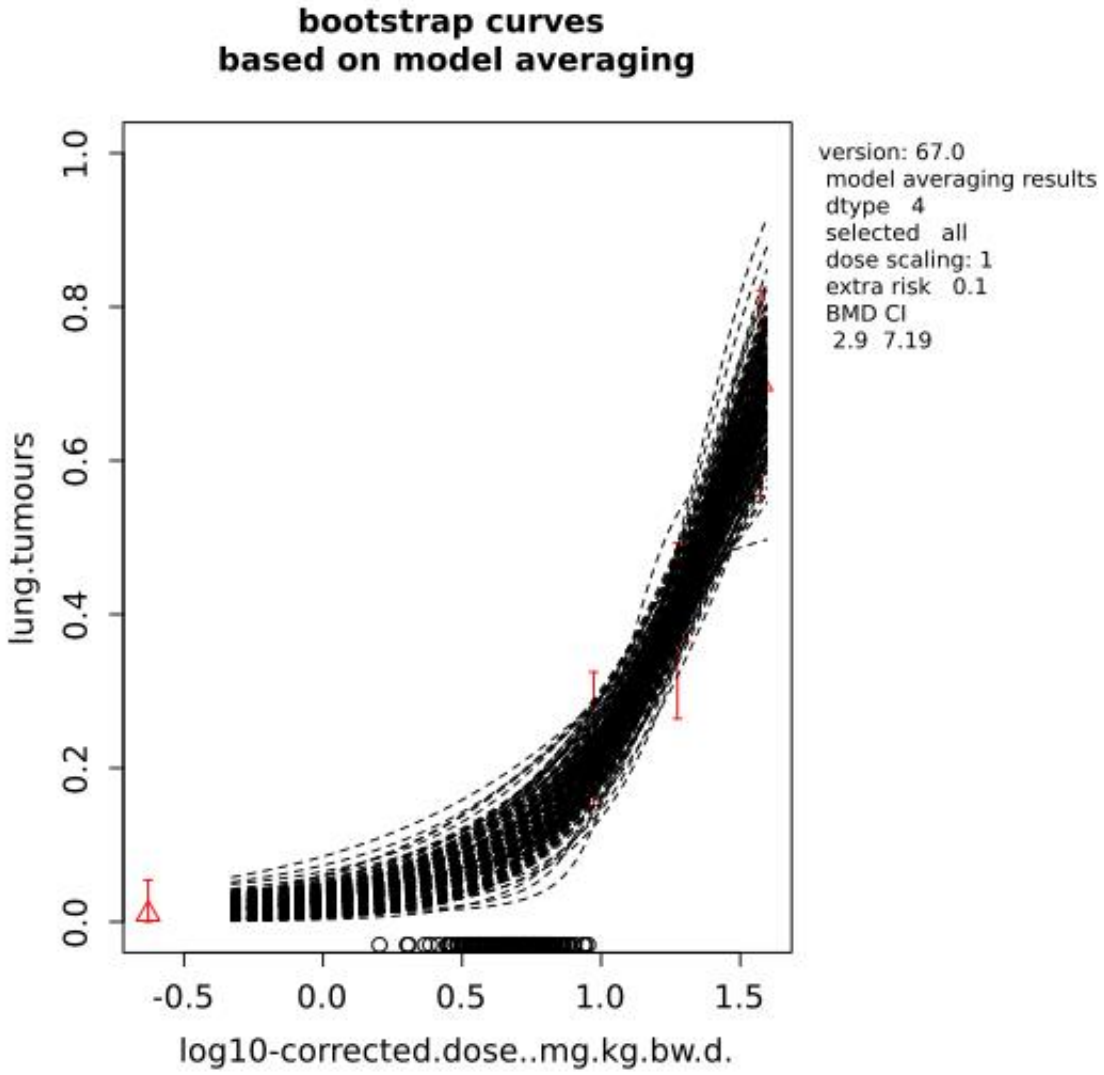
Confidence intervals for the BMD are based on 200 bootstrap data sets.

4,4’-Methylene-bis-[2-chloroaniline] (MOCA)

Visualization



4,4’-Methylene-bis-[2-chloroaniline] (MOCA)



Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for Nitrilotriacetic acid (NTA) and its sodium salts

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: effect.

Data used for analysis:

	dose..mg.kg.bw.d.	effect	n
1	0	0	24
5	0	1	24
2	10	3	23
6	10	1	24
3	100	3	24
7	100	13	24
4	1000	8	24
8	1000	14	24

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is an extra risk of 10% compared to the controls.

When the specified BMR deviates from the default value, the rationale behind the choice made should be described.

The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

Default set of fitted models:

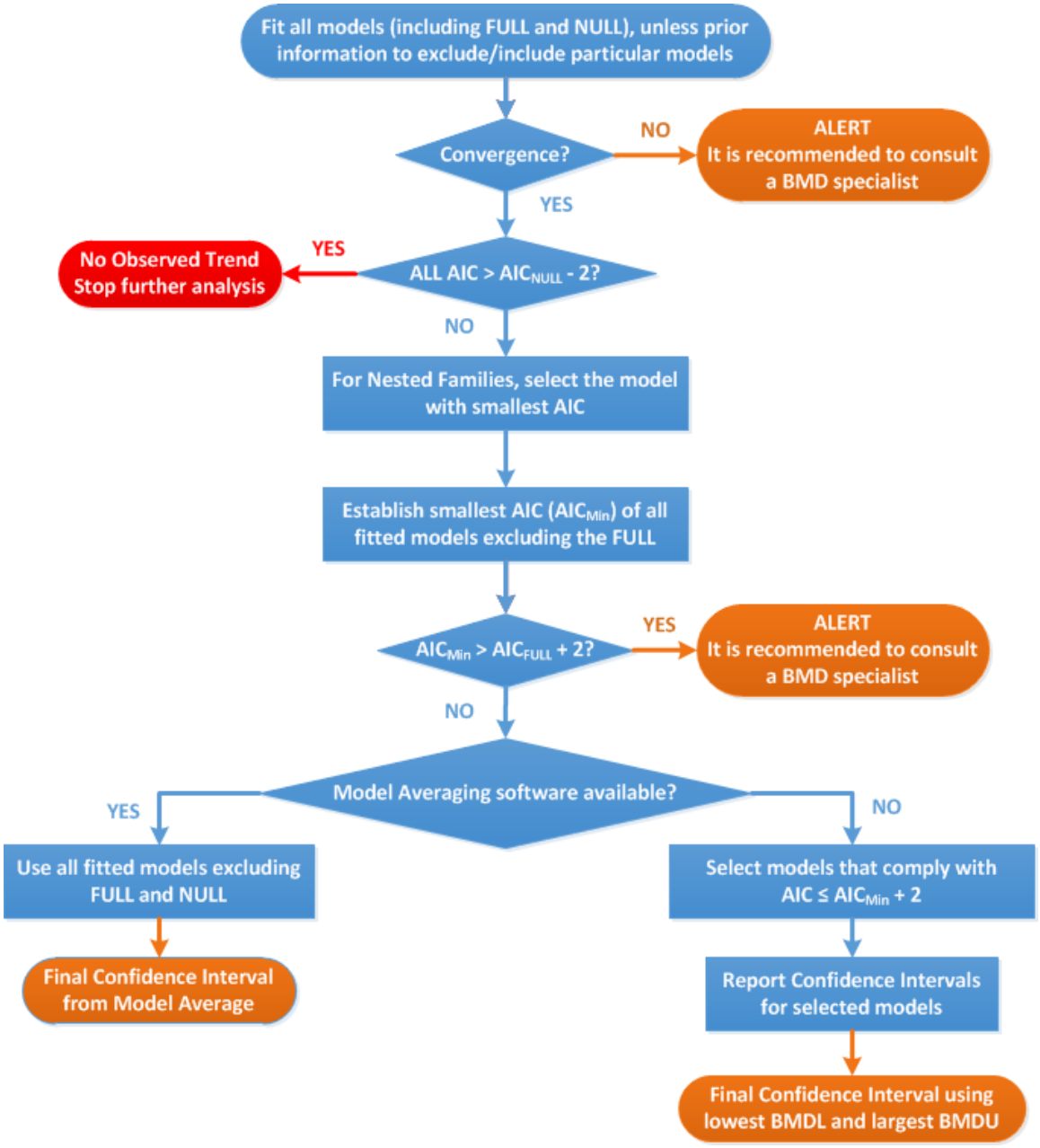
Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean

Logistic	2	$y = \frac{1}{1 + \exp(-a - bx)}$
Probit	2	$y = pnorm((x - a) \cdot b)$
Log-logistic	3	$y = a + \frac{1 - a}{1 + \exp\left(c \cdot \log\left(\frac{b}{x}\right)\right)}$
Log-probit	3	$y = a + (1 - a) \cdot pnorm\left(c \cdot \log\left(\frac{x}{b}\right)\right)$
Weibull	3	$y = a + (1 - a)\left(1 - \exp\left(-\left(\frac{x}{b}\right)^{c}\right)\right)$
Gamma	3	$y = pgamma(bx; c)$
Two-stage	3	$y = a + (1 - a)\left(1 - \exp\left(-\frac{x}{b} - c\left(\frac{x}{b}\right)^{2}\right)\right)$
Exp model 3	3	$y = a \cdot \exp(bx^{d})$
Exp model 5	4	$y = a \cdot (c - (c - 1)\exp(-bx^{d}))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^{d}}{b^{d} + x^{d}}\right)$
Hill model 5	4	$y = a \cdot \left(1 + (c - 1)\frac{x^{d}}{b^{d} + x^{d}}\right)$

For the Exp and Hill family, we fit models with 3 and 4 parameters as listed in the table. The 3-parameter model is selected if the difference in AIC is smaller than 5, otherwise the 4-parameter model is selected.

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: effect

Fitted Models

model	No.par	loglik	AIC	accepted	BMDL	BMDU	BMD	conv
null	1	-101.87	205.74		NA	NA	NA	NA
full	4	-82.20	172.40		NA	NA	NA	NA
two.stage	3	-89.74	185.48	no	NA	NA	171.00	yes
log.logist	3	-83.23	172.46	yes	0.709	34.9	8.45	yes
Weibull	3	-83.43	172.86	yes	0.412	34.0	6.98	yes
log.prob	3	-83.07	172.14	yes	1.100	36.1	9.98	yes
gamma	3	-83.61	173.22	yes	0.223	34.1	5.84	yes
logistic	2	-91.33	186.66	no	NA	NA	342.00	yes
probit	2	-91.76	187.52	no	NA	NA	271.00	no
LVM: Expon. m3-	3	-84.43	174.86	no	NA	NA	18.30	yes
LVM: Hill m3-	3	-83.84	173.68	yes	5.870	36.9	12.80	yes

Estimated Model Parameters

two.stage

estimate for a- : 0.1035

estimate for BMD- : 171.4

estimate for c : 1e-06

log.logist

estimate for a- : 0.01873

estimate for BMD- : 8.447

estimate for c : 0.4451

Weibull

estimate for a- : 0.01877

estimate for BMD- : 6.98

estimate for c : 0.3671

log.prob

estimate for a- : 0.01907

estimate for BMD- : 9.977

estimate for c : 0.2683

gamma

estimate for a- : 0.01891

estimate for BMD- : 5.838

estimate for c : 0.3067

logistic

estimate for a- : -1.875

estimate for BMD- : 341.7

probit

estimate for a- : -1.041

estimate for BMD- : 270.8

EXP

estimate for a- : 1.553

estimate for BMD- : 18.28

estimate for d- : 0.25

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

HILL

estimate for a- : 1.61

estimate for BMD- : 12.76

estimate for d- : 0.25

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

Weights for Model Averaging

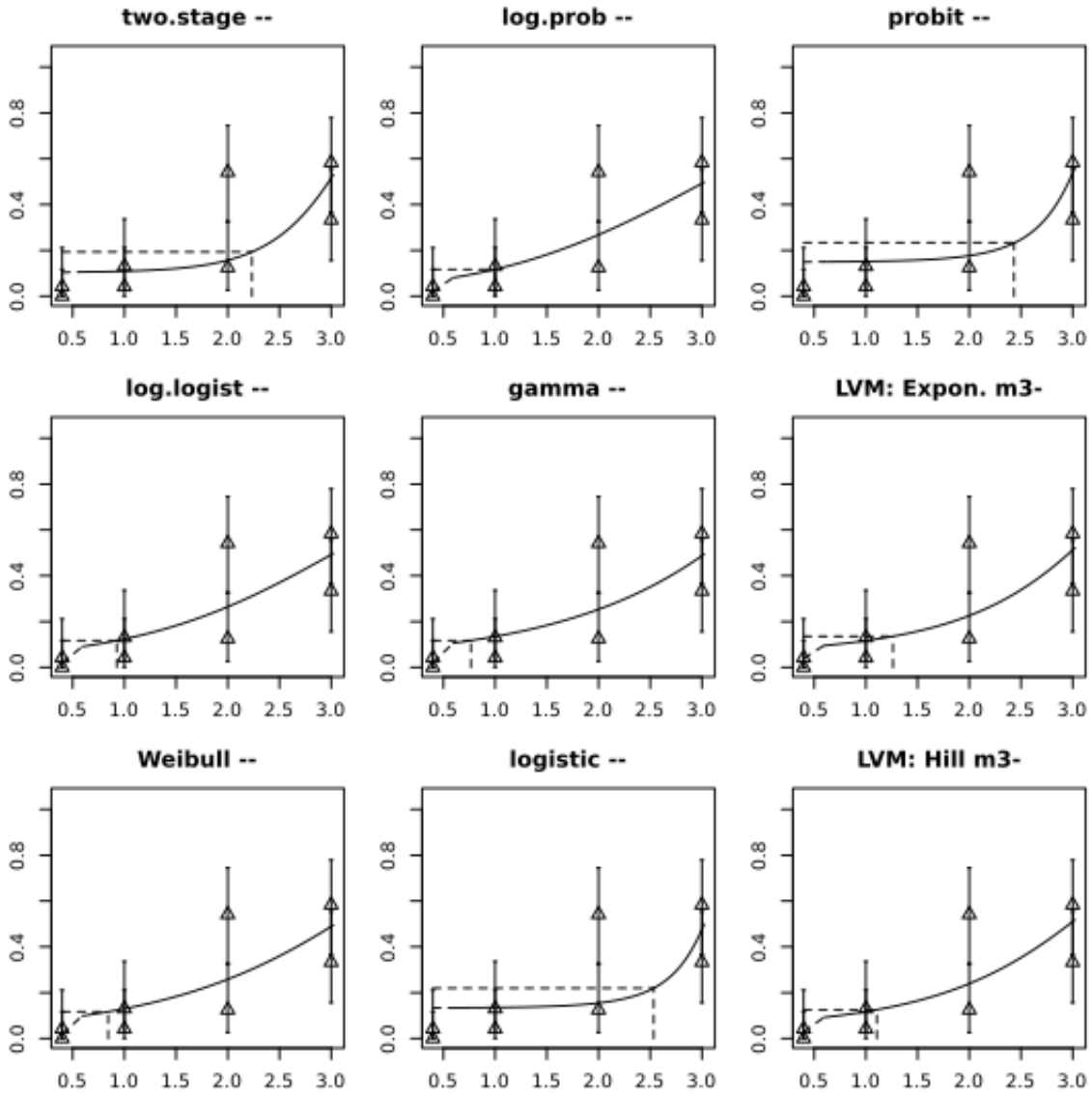
two.stage	log.logist	Weibull	log.prob	gamma	logistic	probit	EXP	HILL
0	0.22	0.18	0.26	0.15	0	0	0.07	0.12

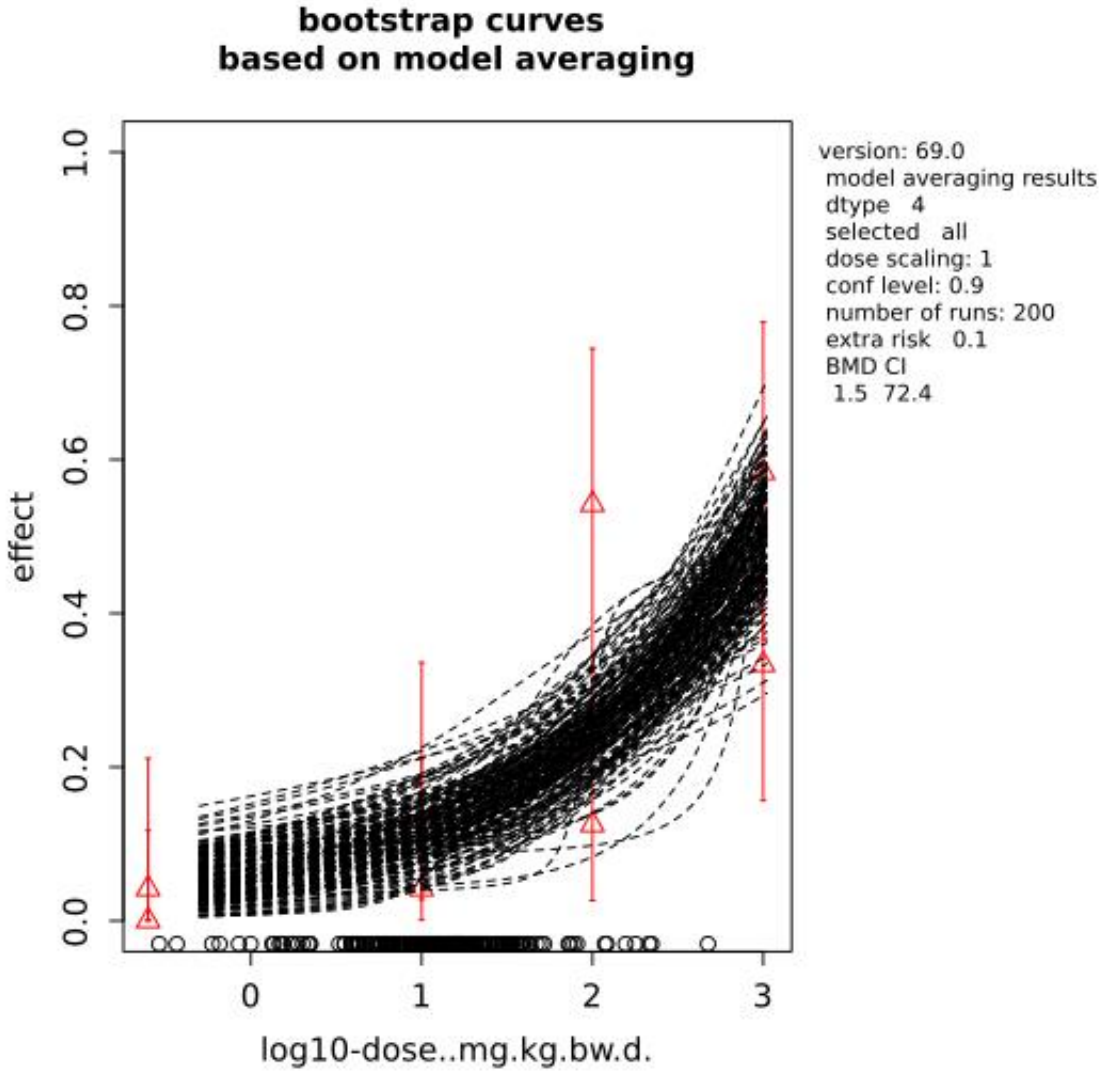
Final BMD Values

subgroup	BMDL	BMDU
all	1.46	72.4

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization





Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for Benzoic acid (no covariate)

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: generalized.inflammation.

Data used for analysis:

	dose..mg.m3.	generalized.inflammation	n
1	0	0	10
5	0	0	10
2	25	3	10
6	25	0	10
3	250	4	10
7	250	5	10
4	1200	8	10
8	1200	9	10

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is an extra risk of 10% compared to the controls.

When the specified BMR deviates from the default value, the rationale behind the choice made should be described.

The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

Default set of fitted models:

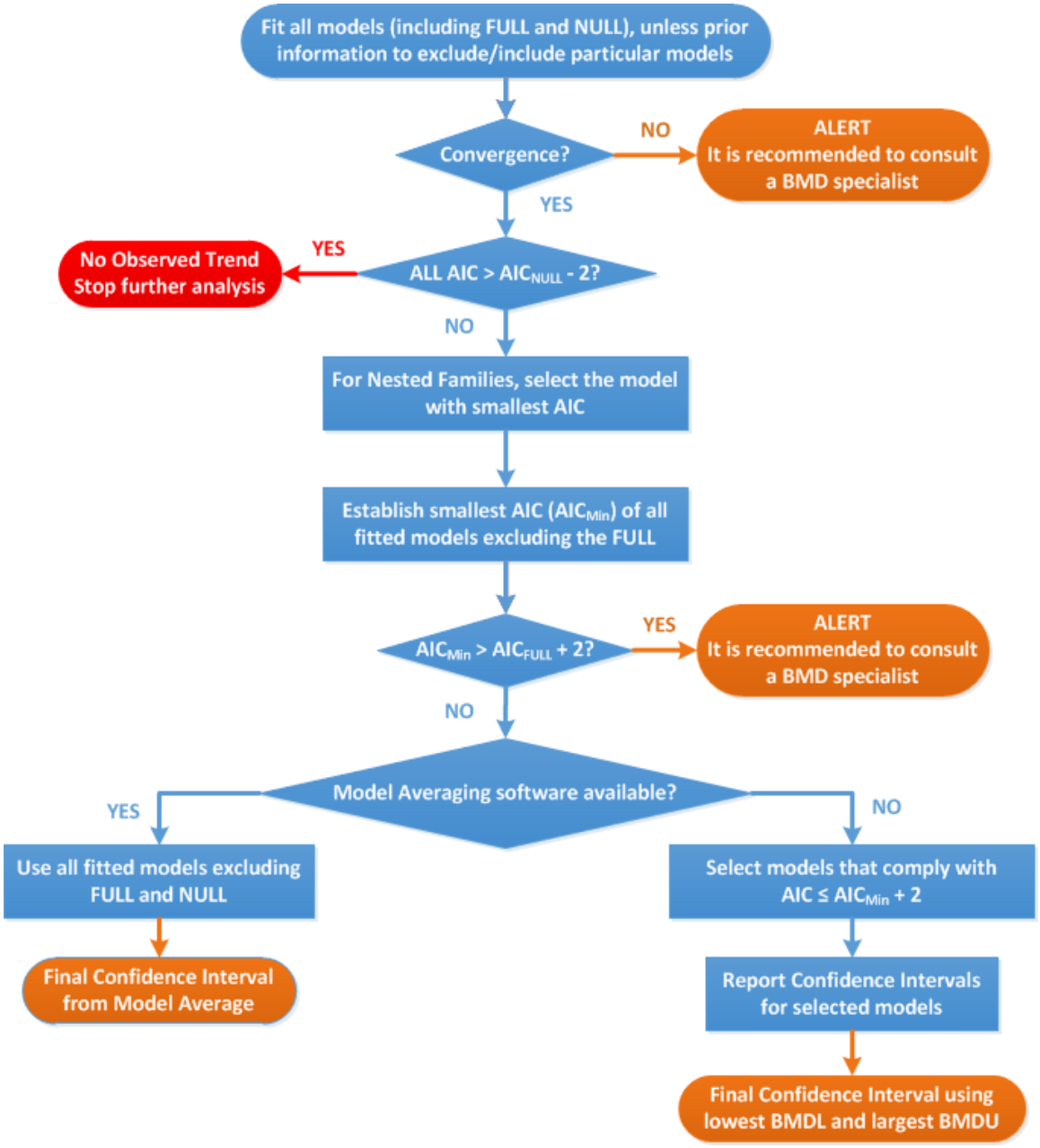
Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean

Logistic	2	$y = \frac{1}{1 + \exp(-a - bx)}$
Probit	2	$y = pnorm((x - a) \cdot b)$
Log-logistic	3	$y = a + \frac{1 - a}{1 + \exp\left(c \cdot \log\left(\frac{b}{x}\right)\right)}$
Log-probit	3	$y = a + (1 - a) \cdot pnorm\left(c \cdot \log\left(\frac{x}{b}\right)\right)$
Weibull	3	$y = a + (1 - a)\left(1 - \exp\left(-\left(\frac{x}{b}\right)^{c}\right)\right)$
Gamma	3	$y = pgamma(bx; c)$
Two-stage	3	$y = a + (1 - a)\left(1 - \exp\left(-\frac{x}{b} - c\left(\frac{x}{b}\right)^{2}\right)\right)$
Exp model 3	3	$y = a \cdot \exp(bx^{d})$
Exp model 5	4	$y = a \cdot (c - (c - 1)\exp(-bx^{d}))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^{d}}{b^{d} + x^{d}}\right)$
Hill model 5	4	$y = a \cdot \left(1 + (c - 1)\frac{x^{d}}{b^{d} + x^{d}}\right)$

For the Exp and Hill family, we fit models with 3 and 4 parameters as listed in the table. The 3-parameter model is selected if the difference in AIC is smaller than 5, otherwise the 4-parameter model is selected.

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: generalized.inflammation

Fitted Models

model	No.par	loglik	AIC	accepted	BMDL	BMDU	BMD	conv
null	1	-52.39	106.78		NA	NA	NA	NA
full	4	-30.67	69.34		NA	NA	NA	NA
two.stage	3	-32.28	70.56	no	NA	NA	57.1	yes
log.logist	3	-31.02	68.04	yes	3.300	51.5	19.2	yes
Weibull	3	-30.74	67.48	yes	1.920	48.5	15.4	yes
log.prob	3	-31.05	68.10	yes	3.880	46.7	19.5	yes
gamma	3	-30.69	67.38	yes	0.837	47.5	12.9	yes
logistic	2	-35.34	74.68	no	NA	NA	180.0	yes
probit	2	-35.32	74.64	no	NA	NA	183.0	yes
LVM: Expon. m3-	3	-30.85	67.70	yes	4.080	54.4	17.5	yes
LVM: Hill m3-	3	-30.84	67.68	yes	3.090	58.1	17.3	yes

Estimated Model Parameters

two.stage

estimate for a- : 0.03544

estimate for BMD- : 57.13

estimate for c : 1e-06

log.logist

estimate for a- : 1e-06

estimate for BMD- : 19.22

estimate for c : 0.8826

Weibull

estimate for a- : 1e-06

estimate for BMD- : 15.41

estimate for c : 0.6549

log.prob

estimate for a- : 1e-06

estimate for BMD- : 19.48

estimate for c : 0.5247

gamma

estimate for a- : 1e-06

estimate for BMD- : 12.88

estimate for c : 0.5433

logistic

estimate for a- : -1.803

estimate for BMD- : 179.8

probit

estimate for a- : -1.046

estimate for BMD- : 183.4

EXP

estimate for a- : 1.853

estimate for BMD- : 17.53

estimate for d- : 0.25

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

HILL

estimate for a- : 2.107

estimate for BMD- : 17.26

estimate for d- : 0.2707

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

Weights for Model Averaging

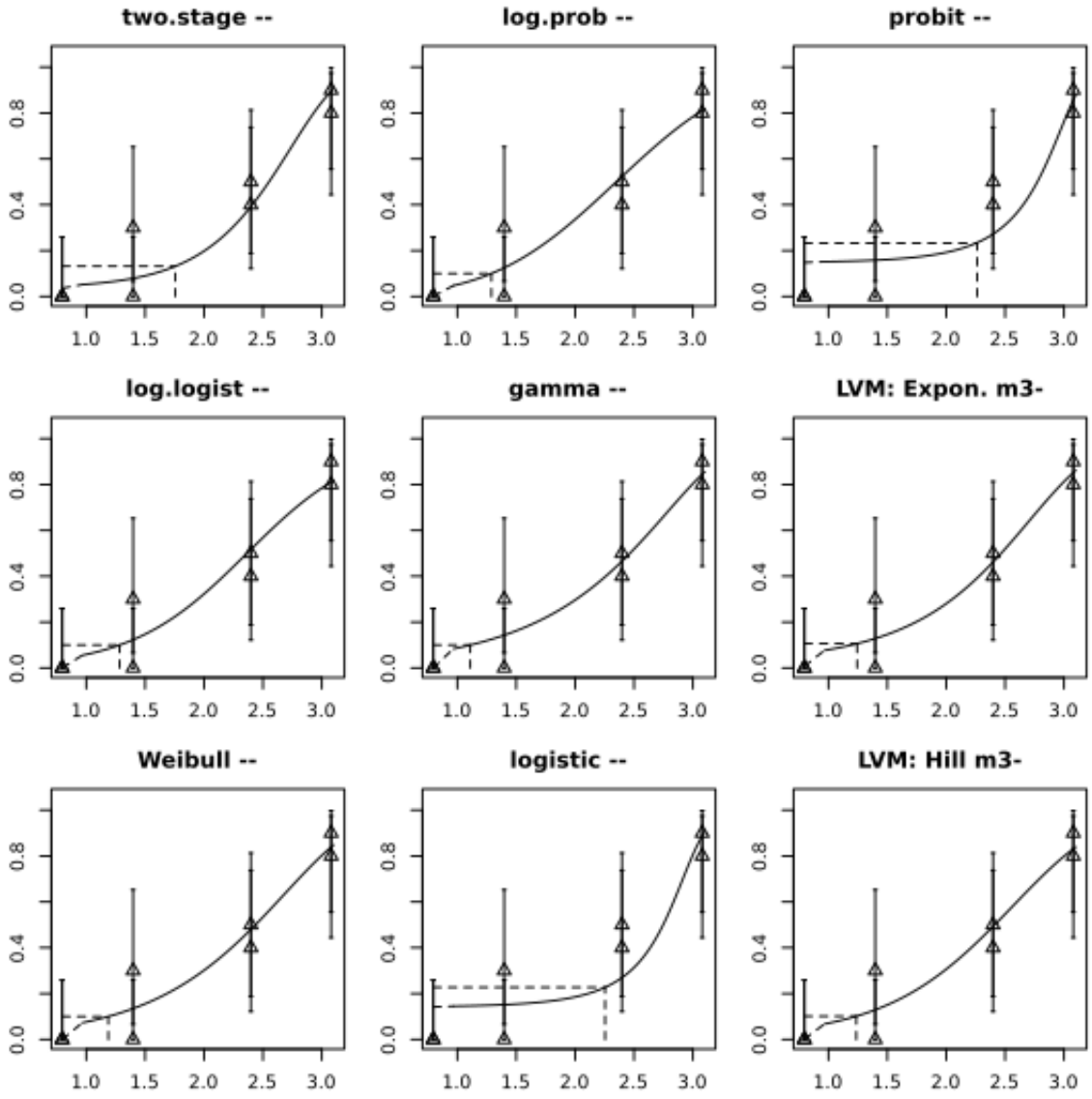
two.stage	log.logist	Weibull	log.prob	gamma	logistic	probit	EXP	HILL
0.04	0.13	0.18	0.13	0.19	0	0	0.16	0.16

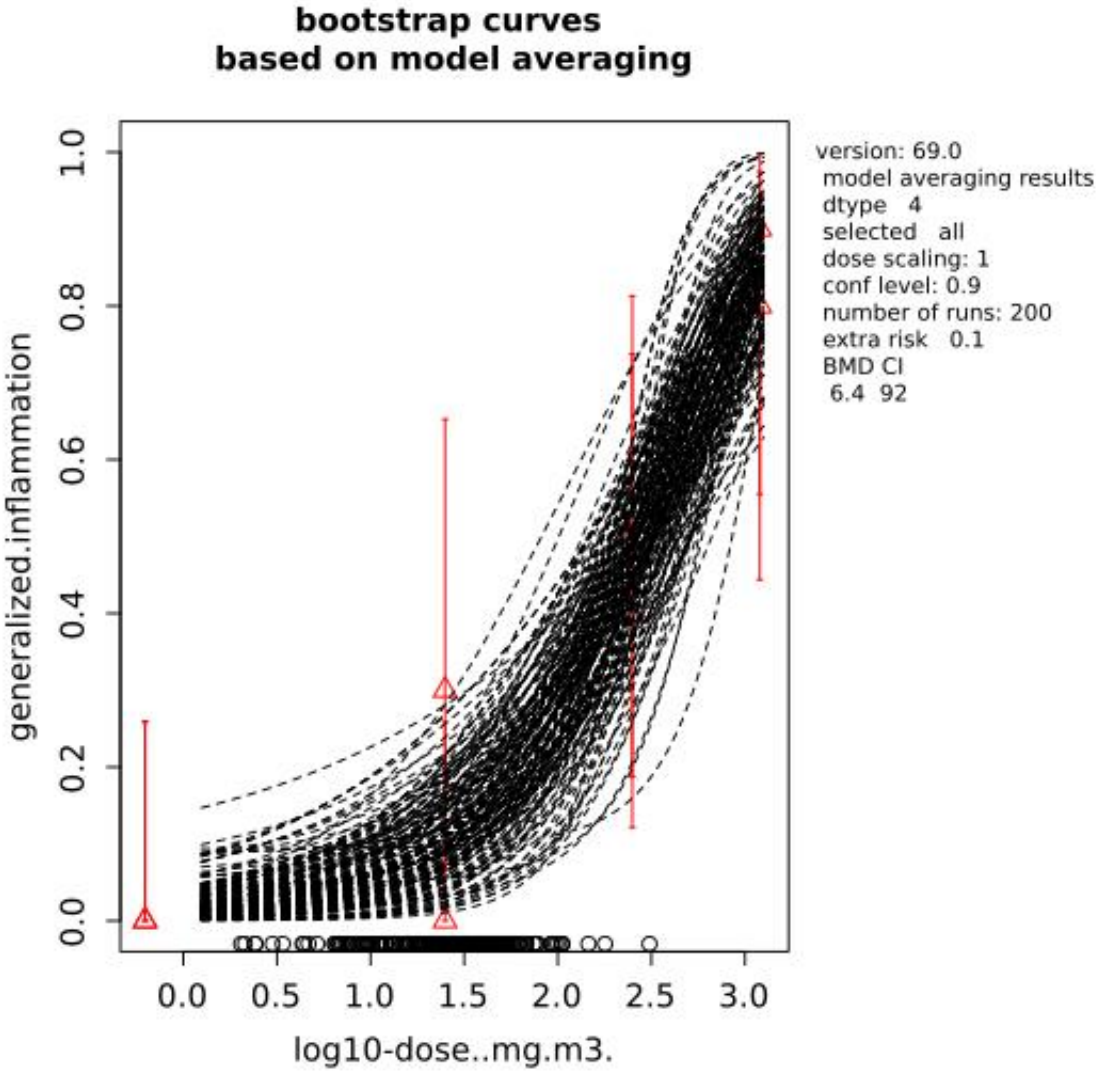
Final BMD Values

subgroup	BMDL	BMDU
all	6.36	92

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization





Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for Nalidixic acid (female animals)

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: bw.

Warning: You selected 6 rows to be excluded from the analysis.

concentration	n	bw	SEM	sex
0	10	36.1	0.89	m
1000	10	35.0	0.64	m
2000	10	34.9	0.71	m
4000	10	33.6	0.41	m
8000	10	32.4	0.47	m
16000	10	31.4	0.71	m

Data used for analysis:

	concentration	bw	SEM	n
7	0	29.4	1.13	10
8	1000	28.2	1.18	10

9	2000	28.7	1.06	9
10	4000	27.1	0.42	10
11	8000	24.8	0.84	10
12	16000	23.6	0.55	10

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is a 10% change in mean response compared to the controls. The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

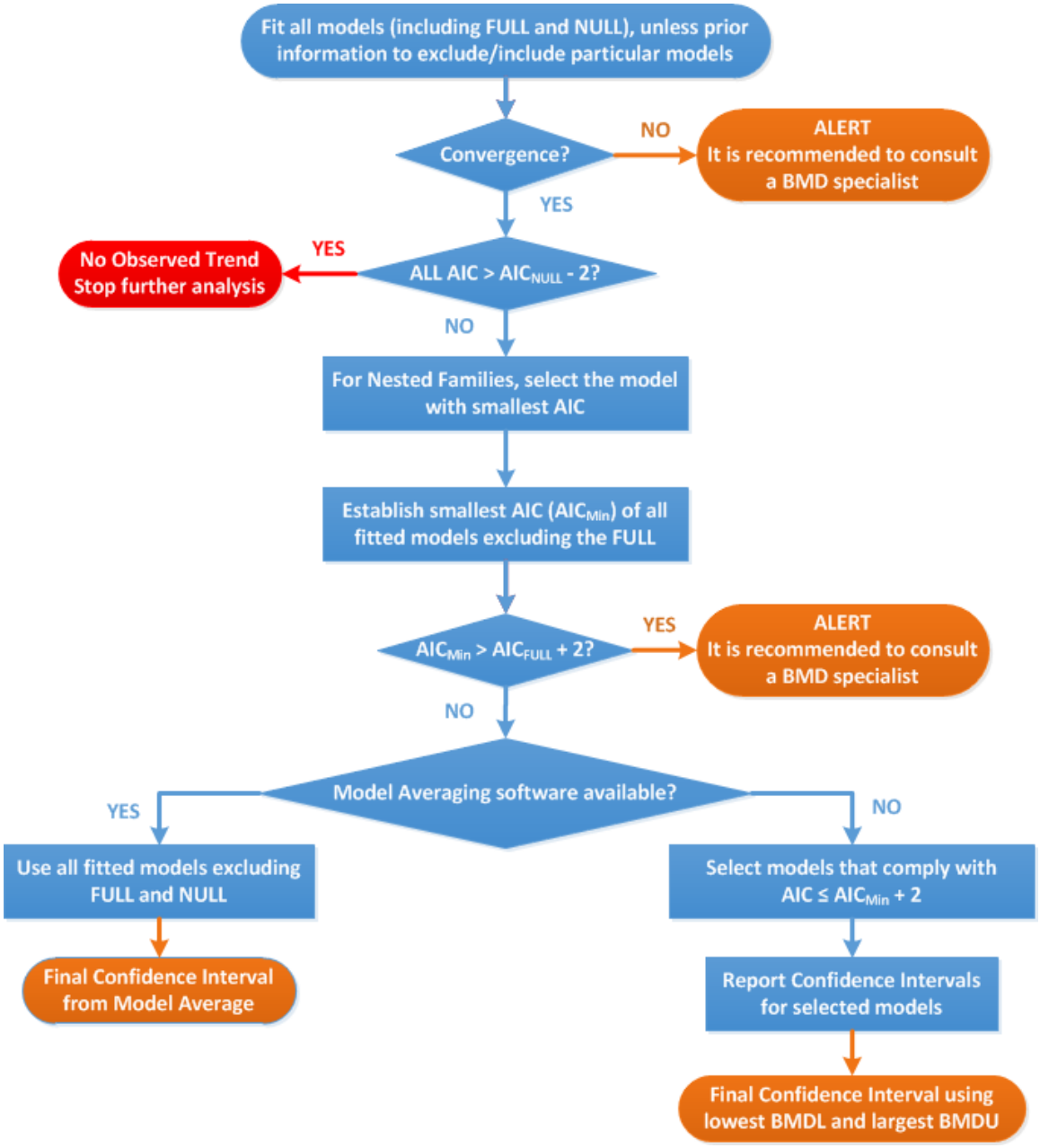
Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

Default set of fitted models:

Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean
Exp model 3	3	$y = a \cdot \exp(bx^d)$
Exp model 4	4	$y = a \cdot (c - (c-1)\exp(-bx^d))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^d}{b^d + x^d}\right)$
Hill model 4	4	$y = a \cdot \left(1 - \frac{(c-1) \cdot x^d}{b^d + x^d}\right)$
Inverse Exponential	4	$y = a \cdot (1 + (c-1)\exp(-bx^{-d}))$
Log-Normal Family	4	$y = a \cdot (1 + (c-1)\Phi(\ln b + d\ln x))$

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: bw

Fitted Models

model	converged	loglik	npar	AIC
full model	yes	53.74	7	-93.48
null model	yes	38.94	2	-73.88
Expon. m3-	yes	52.71	4	-97.42
Expon. m5-	yes	53.06	5	-96.12
Hill m3-	yes	52.71	4	-97.42
Hill m5-	yes	53.16	5	-96.32
Inv.Expon. m3-	yes	52.80	4	-97.60
Inv.Expon. m5-	yes	53.18	5	-96.36
LN m3-	yes	52.76	4	-97.52
LN m5-	yes	53.09	5	-96.18

Estimated Model Parameters

EXP

estimate for var- : 0.009807

estimate for a- : 29.29

estimate for CED- : 5160

estimate for d- : 0.6878

HILL

estimate for var- : 0.009807

estimate for a- : 29.29

estimate for CED- : 5160

estimate for d- : 0.6891

INVEXP

estimate for var- : 0.009778

estimate for a- : 29.21

estimate for CED- : 5240

estimate for d- : 0.1222

LOGN

estimate for var- : 0.00979

estimate for a- : 29.24

estimate for CED- : 5204

estimate for d- : 0.2296

Weights for Model Averaging

EXP	HILL	INVEXP	LOGN
0.24	0.24	0.26	0.25

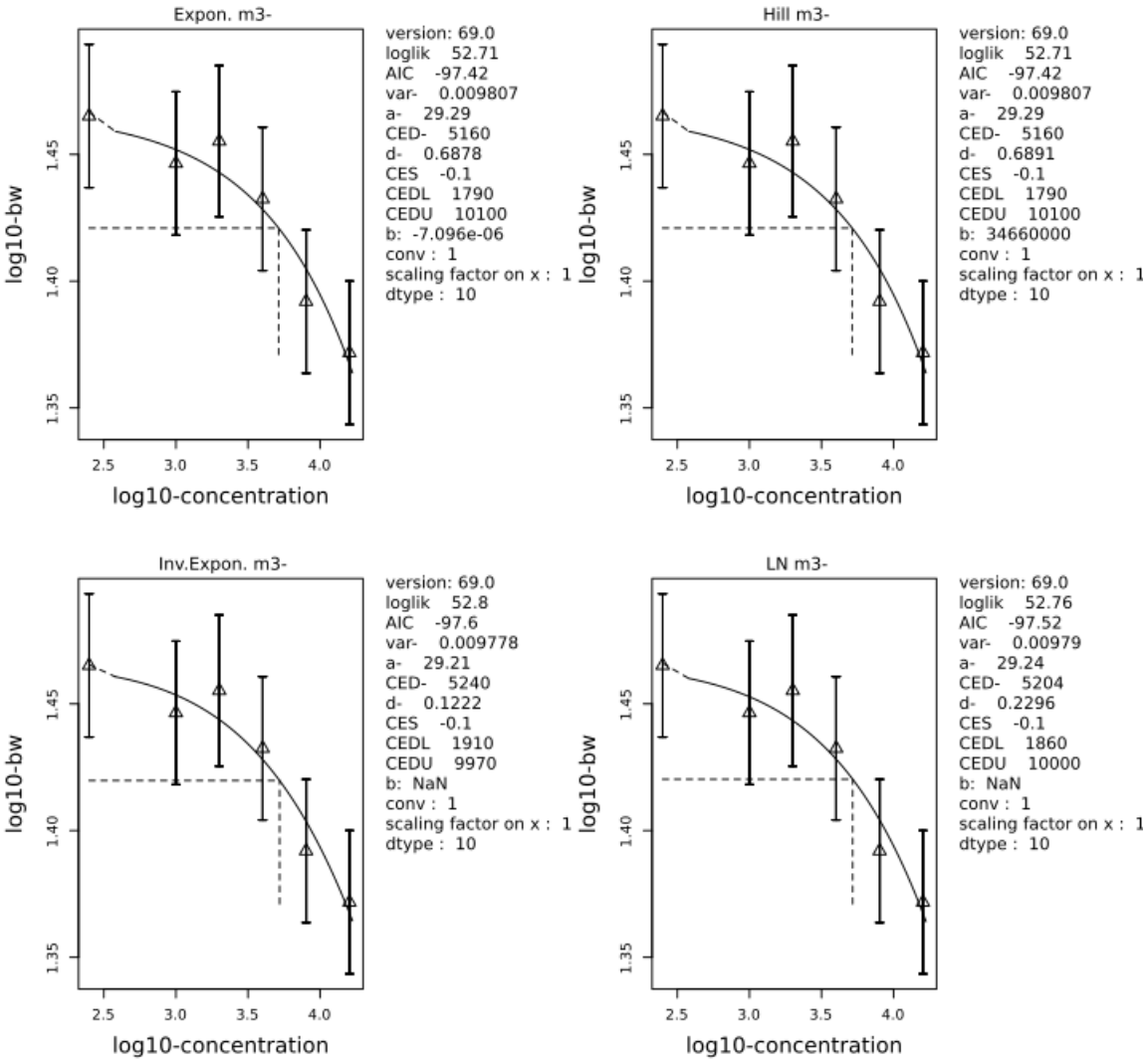
Final BMD Values

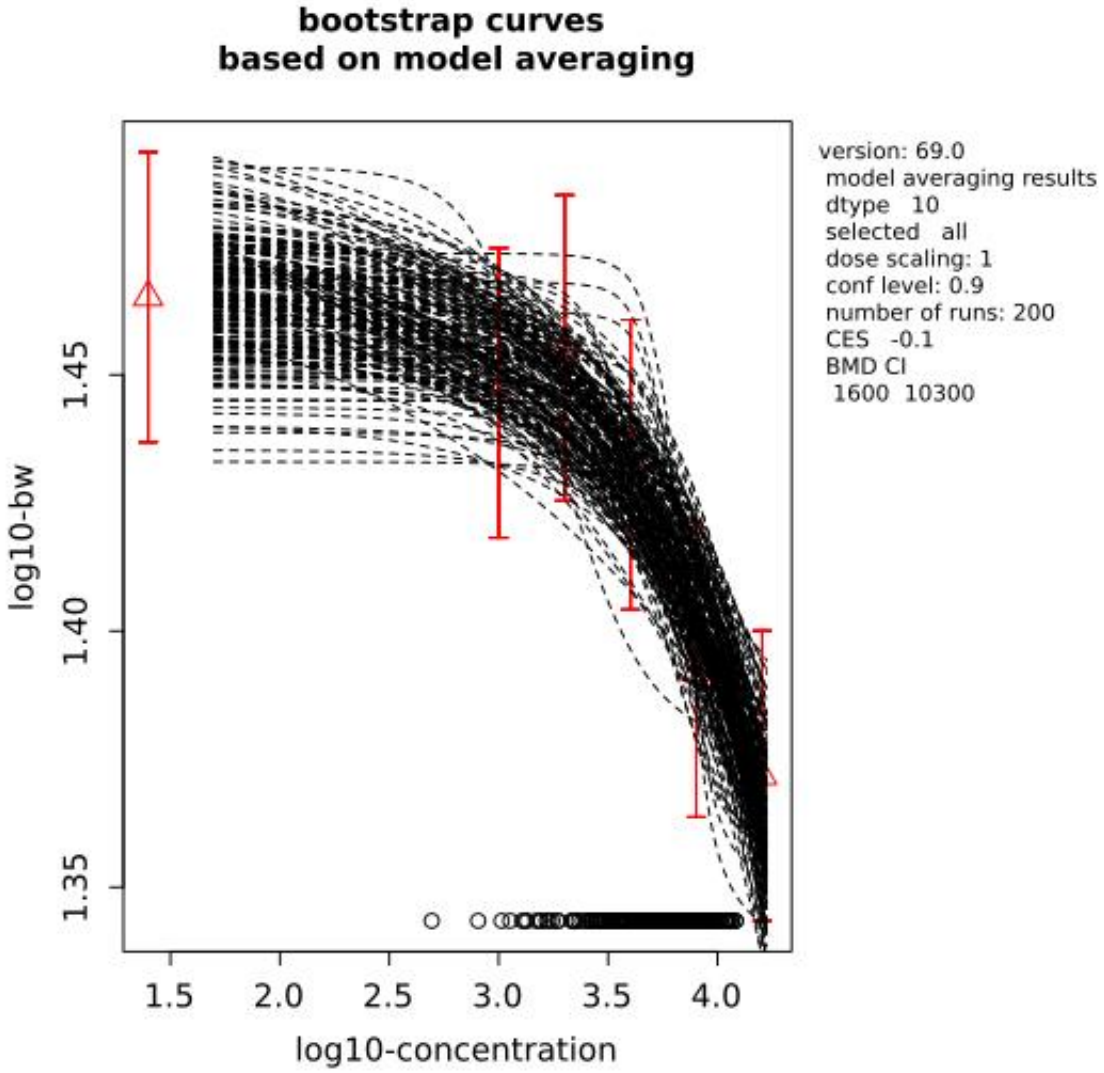
endpoint	subgroup	BMDL	BMDU
bw	all	1650	10300

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization

Nalidixic acid (females)





Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for Nalidixic acid (males)

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: bw.

Warning: You selected 6 rows to be excluded from the analysis.

	concentration	n	bw	SEM	sex
7	0	10	29.4	1.13	f
8	1000	10	28.2	1.18	f
9	2000	9	28.7	1.06	f
10	4000	10	27.1	0.42	f
11	8000	10	24.8	0.84	f
12	16000	10	23.6	0.55	f

Data used for analysis:

concentration	bw	SEM	n
0	36.1	0.89	10
1000	35.0	0.64	10

2000	34.9	0.71	10
4000	33.6	0.41	10
8000	32.4	0.47	10
16000	31.4	0.71	10

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is a 10% change in mean response compared to the controls. The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

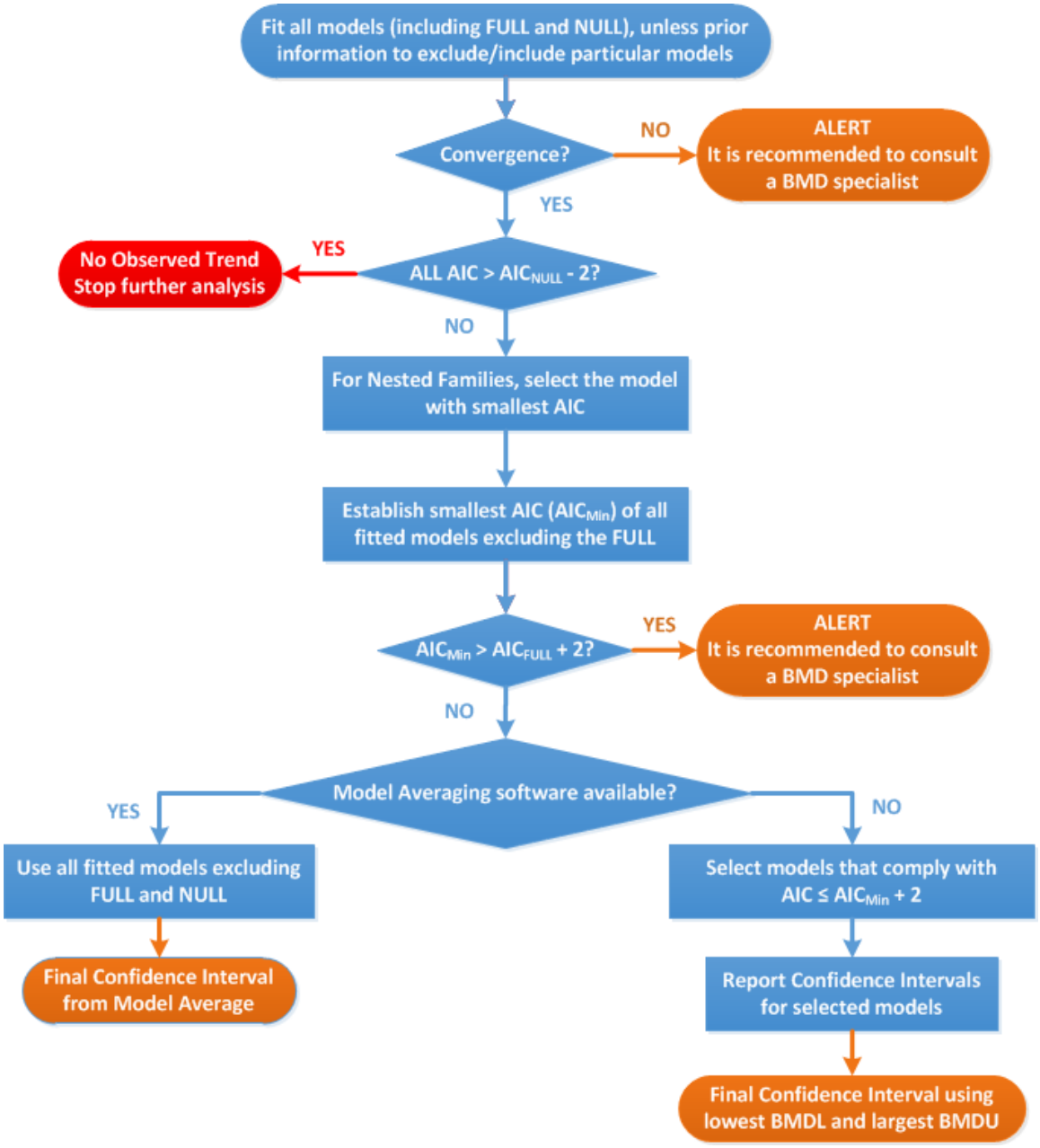
Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

Default set of fitted models:

Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean
Exp model 3	3	$y = a \cdot \exp(bx^d)$
Exp model 4	4	$y = a \cdot (c - (c-1)\exp(-bx^d))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^d}{b^d + x^d}\right)$
Hill model 4	4	$y = a \cdot \left(1 - \frac{(c-1) \cdot x^d}{b^d + x^d}\right)$
Inverse Exponential	4	$y = a \cdot (1 + (c-1)\exp(-bx^{-d}))$
Log-Normal Family	4	$y = a \cdot (1 + (c-1)\Phi(\ln b + d\ln x))$

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: bw

Fitted Models

model	converged	loglik	npar	AIC
full model	yes	85.98	7	-157.96
null model	yes	70.35	2	-136.70
Expon. m3-	yes	85.62	4	-163.24
Expon. m5-	yes	85.73	5	-161.46
Hill m3-	yes	85.62	4	-163.24
Hill m5-	yes	85.79	5	-161.58
Inv.Expon. m3-	yes	85.69	4	-163.38
Inv.Expon. m5-	yes	85.76	5	-161.52
LN m3-	yes	85.66	4	-163.32
LN m5-	yes	85.78	5	-161.56

Estimated Model Parameters

EXP

estimate for var- : 0.003374

estimate for a- : 36.06

estimate for CED- : 8923

estimate for d- : 0.5568

HILL

estimate for var- : 0.003374

estimate for a- : 36.06

estimate for CED- : 8922

estimate for d- : 0.5574

INVEXP

estimate for var- : 0.003366

estimate for a- : 36.03

estimate for CED- : 8842

estimate for d- : 0.09109

LOGN

estimate for var- : 0.003369

estimate for a- : 36.05

estimate for CED- : 8875

estimate for d- : 0.1774

Weights for Model Averaging

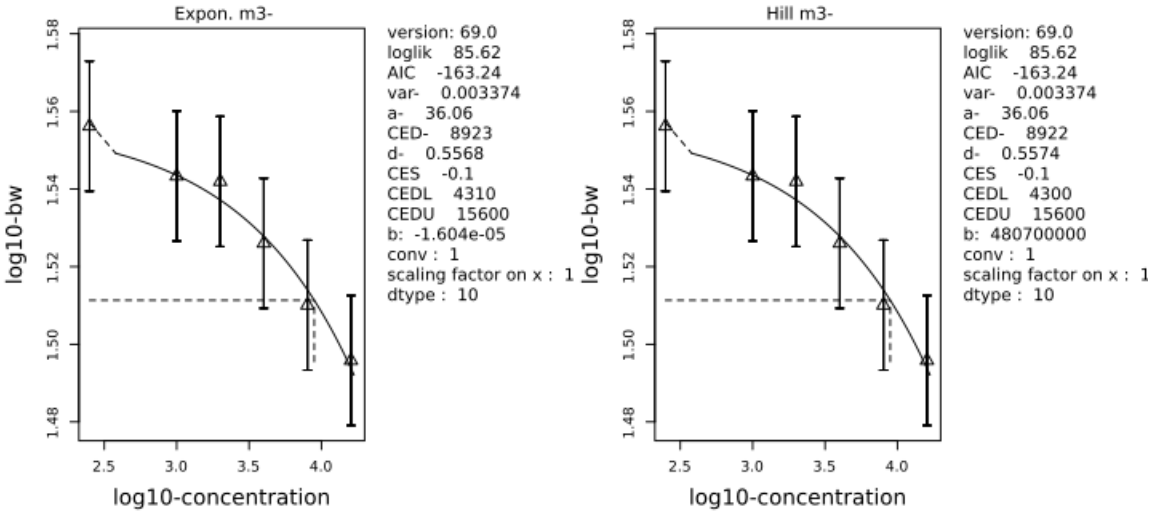
EXP	HILL	INVEXP	LOGN
0.24	0.24	0.26	0.25

Final BMD Values

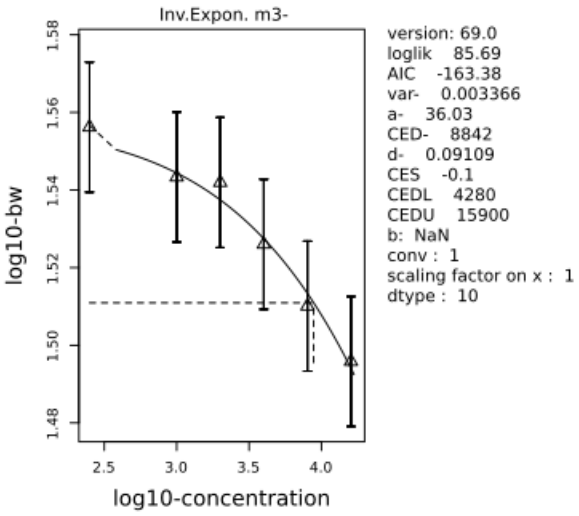
endpoint	subgroup	BMDL	BMDU
bw	all	4410	14800

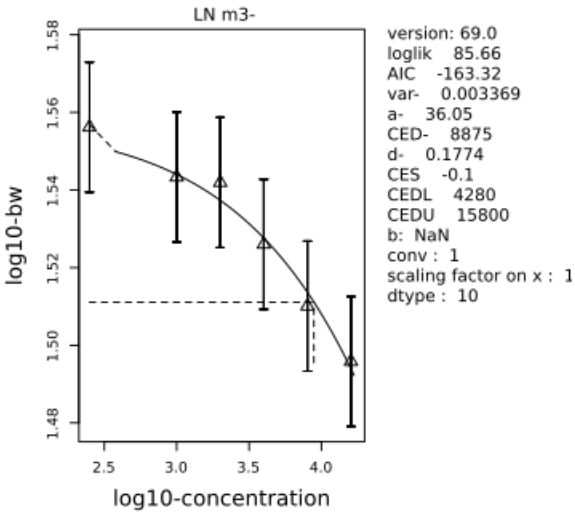
Confidence intervals for the BMD are based on 200 bootstrap data sets.

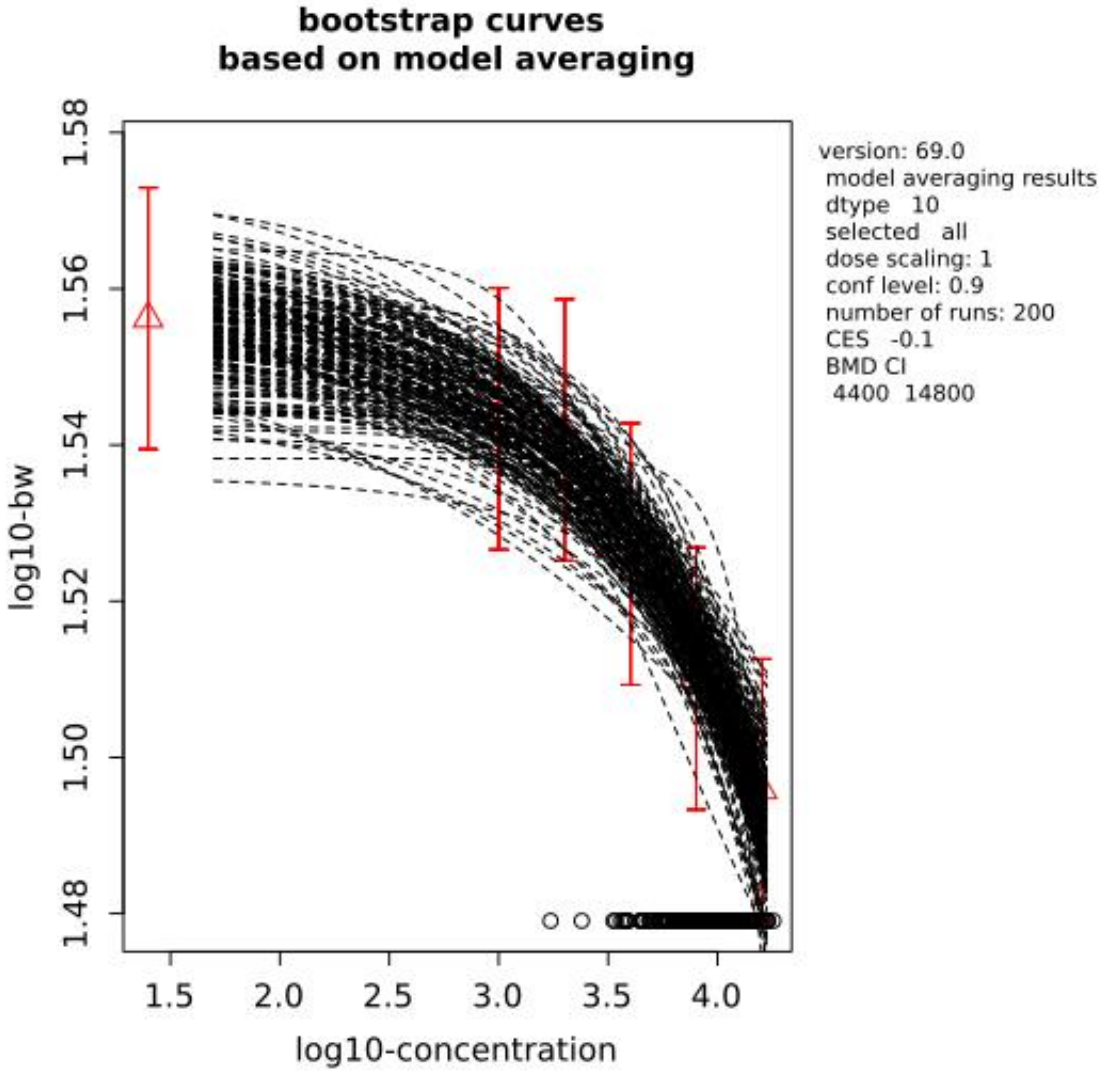
Visualization



Nalidixic acid (males)







Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for 1,1,2,2 Tetrachloroethane

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: relative.liver.weight..mean..

Data used for analysis:

	dose..mg.kg.bw.d.	relative.liver. weight..mean.	relative.liver. weight..standard.error.	n
1	0	34.79	0.42	10
7	0	35.07	0.56	10
2	20	36.72	0.44	10
8	20	36.69	0.36	10
3	40	41.03	0.85	10
9	40	37.84	0.51	10
4	80	45.61	0.52	10
10	80	44.20	0.27	10
5	170	44.68	0.45	10
11	170	48.03	0.89	10

6	320	52.23	1.42	10
12	320	58.40	1.42	10

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is a 7% change in mean response compared to the controls. The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

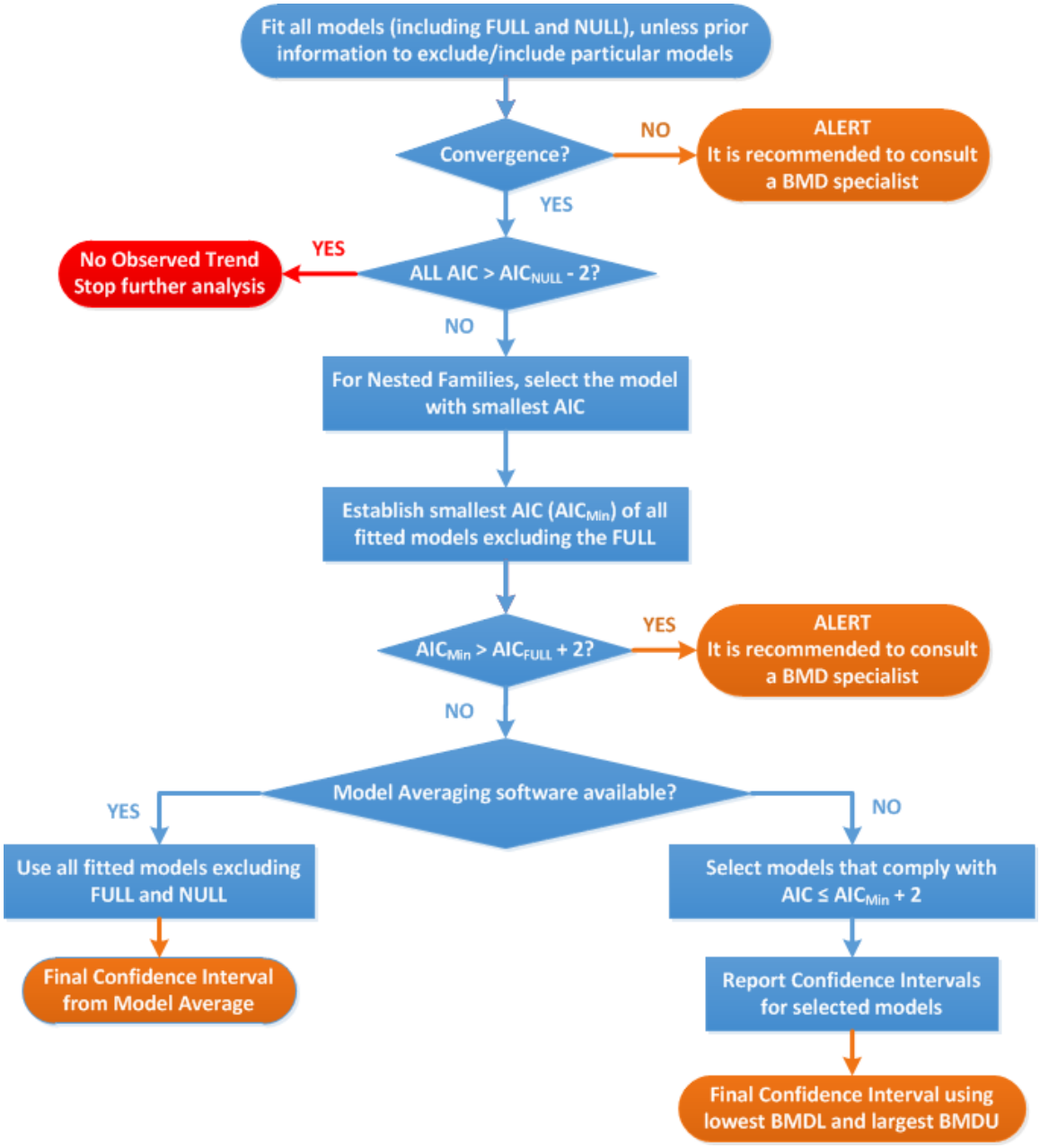
Default set of fitted models:

Model	Number of parameters	Formula
Null	1	$y = a$

Full	no. of groups	$y = \text{group mean}$
Exp model 3	3	$y = a \cdot \exp(bx^d)$
Exp model 4	4	$y = a \cdot (c - (c-1)\exp(-bx^d))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^d}{b^d + x^d}\right)$
Hill model 4	4	$y = a \cdot \left(1 - \frac{(c-1) \cdot x^d}{b^d + x^d}\right)$
Inverse Exponential	4	$y = a \cdot (1 + (c-1)\exp(-bx^{-d}))$
Log-Normal Family	4	$y = a \cdot (1 + (c-1)\Phi(\ln b + d\ln x))$

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: relative.liver.weight..mean.

Fitted Models

model	converged	loglik	npar	AIC
full model	yes	191.91	13	-357.82
null model	yes	45.66	2	-87.32
Expon. m3-	yes	157.07	4	-306.14
Expon. m5-	yes	157.68	5	-305.36
Hill m3-	yes	157.08	4	-306.16
Hill m5-	yes	157.90	5	-305.80
Inv.Expon. m3-	yes	158.17	4	-308.34
Inv.Expon. m5-	yes	158.90	5	-307.80
LN m3-	yes	157.73	4	-307.46
LN m5-	yes	158.55	5	-307.10

Attention: the AIC of the best model (minimum AIC) is more than two units larger than that of the full model. This might indicate a problem in the data, in particular when the difference is much larger than two units (e.g. > 5). You might check the following options:1. In real-life studies, not all experimental factors are completely randomized over all animals (experimental units), e.g. animals were housed in the same cage within a given dose group, or order of treatments were not randomized over individual animals. Another option is that individual outlying animals distort the mean response of one or more treatment groups. This may lead to fluctuations in the (mean) responses among treatment groups that are larger than expected from random sampling error, resulting in an AIC difference with the full model larger than 2 units.2. the data consist of subgroups not taken into account in the model (e.g. various studies, or two sexes) 3. the data contain litter effects not taken into account 4. the response in the top dose group deviates substantially from the fitted model (check the CI around the observed (mean) response); Associated actions for each of these four options are:1. the greater scatter in (mean) responses will result in a wider BMD CI; normally, no further action is needed, as the BMD approach is relatively robust to such devations. You might check this by leaving out specific treatment groups (one by one) and check if this has a major impact on the BMD CI.2.

use the factor defining the subgroups as a covariate and re-analyse the data 3. re-analyse the data with litter effects taken into account 4. consider to leave out the top dose; it is not recommended to leave out two high dose groups.

Estimated Model Parameters

EXP

estimate for var- : 0.004272

estimate for a- : 34.59

estimate for CED- : 12.15

estimate for d- : 0.5886

HILL

estimate for var- : 0.004271

estimate for a- : 34.59

estimate for CED- : 12.19

estimate for d- : 0.5905

INVEXP

estimate for var- : 0.004194

estimate for a- : 34.67

estimate for CED- : 15.49

estimate for d- : 0.1165

LOGN

estimate for var- : 0.004225

estimate for a- : 34.63

estimate for CED- : 14.07

estimate for d- : 0.2091

Weights for Model Averaging

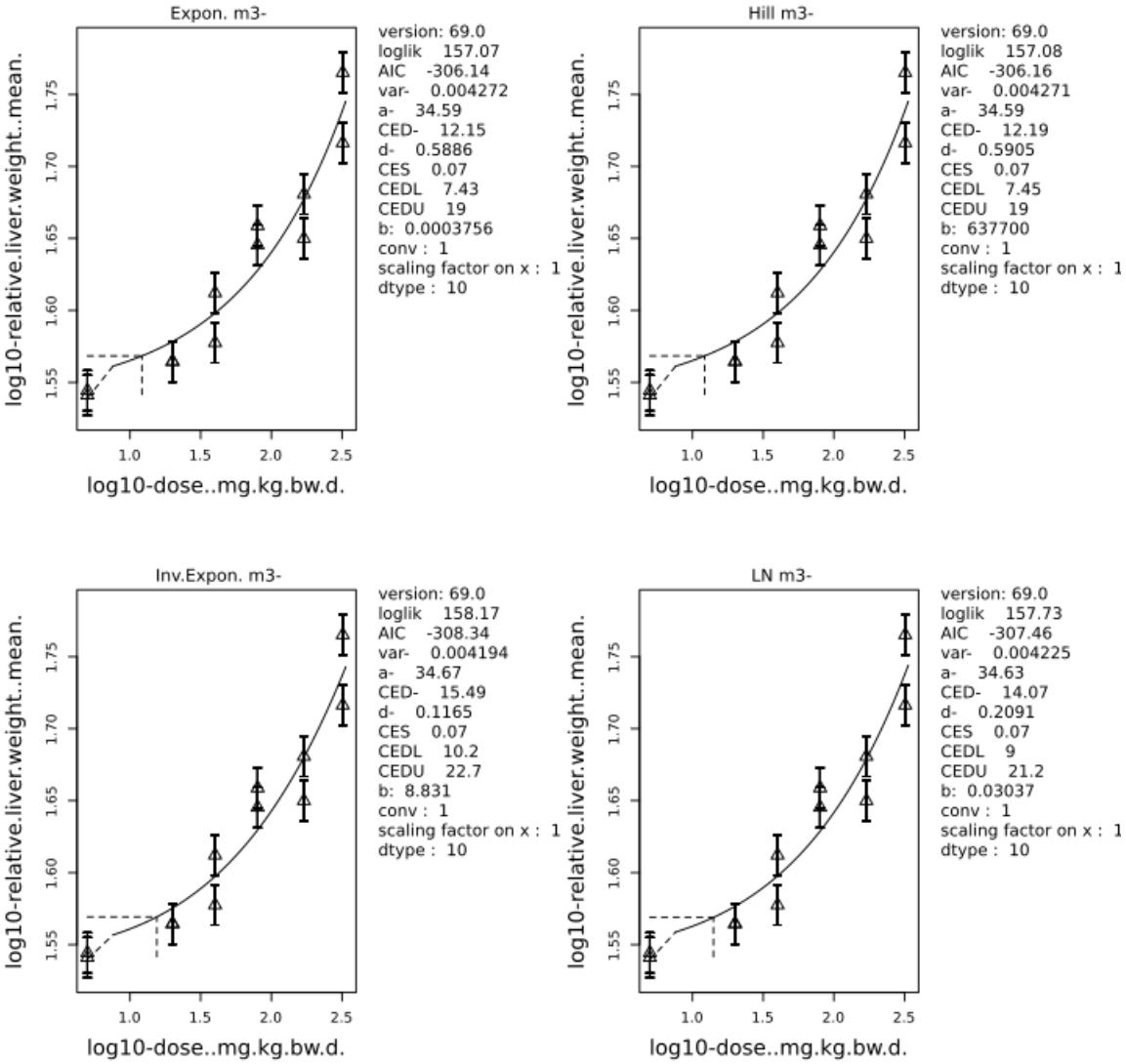
EXP	HILL	INVEXP	LOGN
0.14	0.15	0.43	0.28

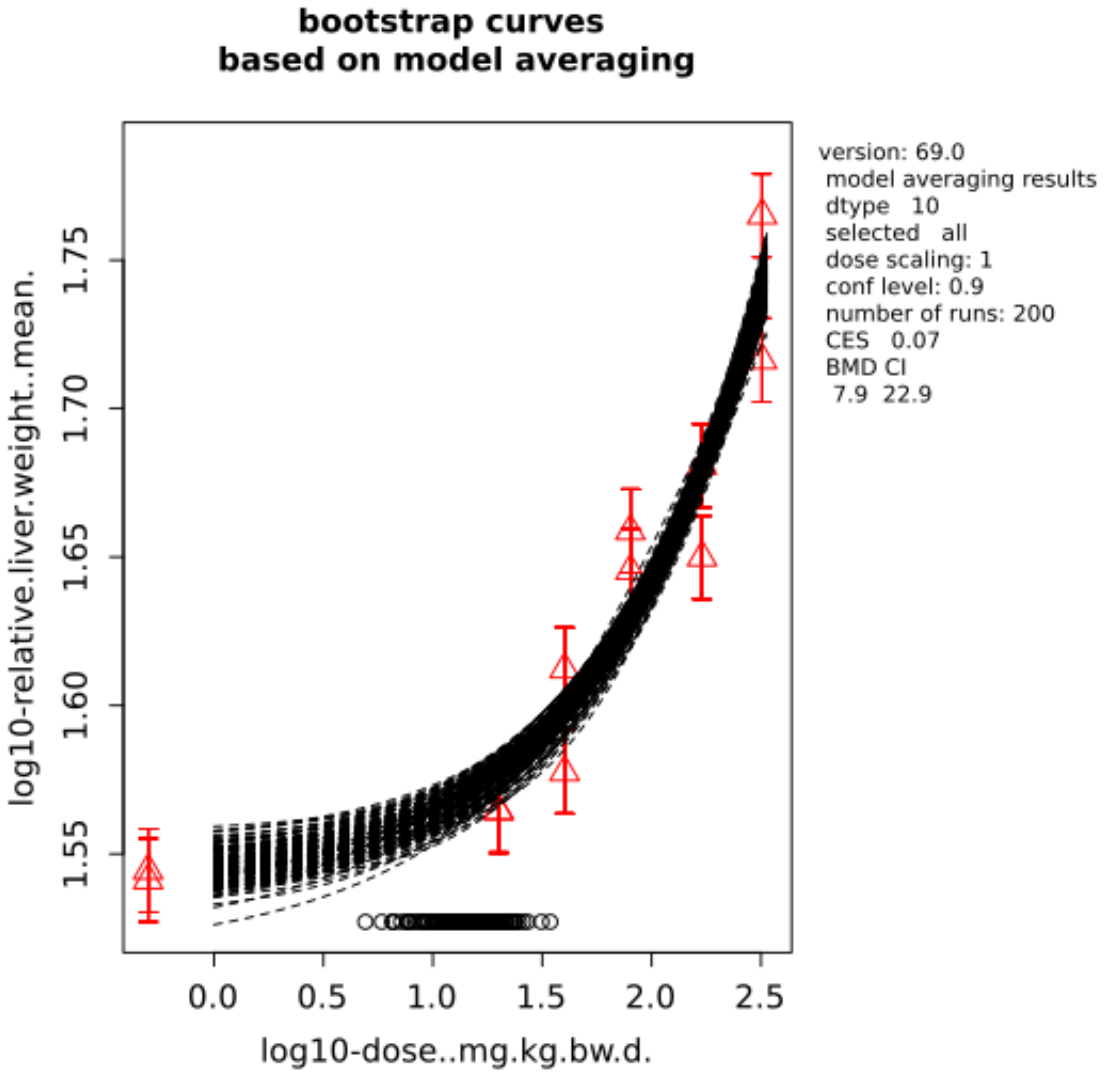
Final BMD Values

endpoint	subgroup	BMDL	BMDU
relative.liver.weight..mean.	all	7.87	22.9

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization





Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: relative.liver.weight..mean..

Data used for analysis:

	dose..mg.kg.bw.d.	relative.liver.weight..mean.	relative.liver.weight..standard.error.	n
1	0	34.79	0.42	10
7	0	35.07	0.56	10
2	20	36.72	0.44	10
8	20	36.69	0.36	10
3	40	41.03	0.85	10
9	40	37.84	0.51	10
4	80	45.61	0.52	10
10	80	44.20	0.27	10
5	170	44.68	0.45	10
11	170	48.03	0.89	10

6	320	52.23	1.42	10
12	320	58.40	1.42	10

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is a 20% change in mean response compared to the controls. The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

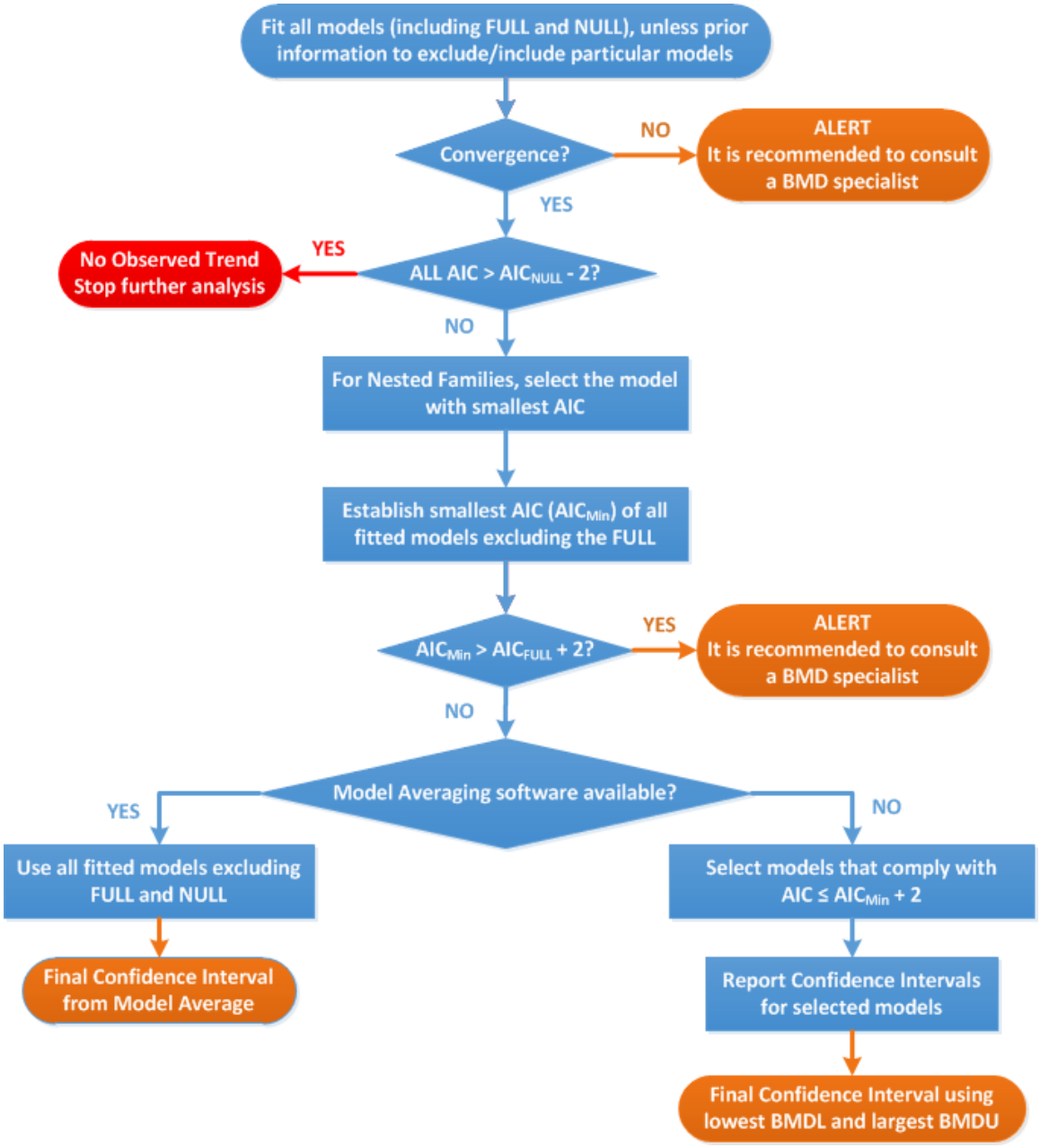
Default set of fitted models:

Model	Number of parameters	Formula
Null	1	$y = a$

Full	no. of groups	$y = \text{group mean}$
Exp model 3	3	$y = a \cdot \exp(bx^d)$
Exp model 4	4	$y = a \cdot (c - (c-1)\exp(-bx^d))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^d}{b^d + x^d}\right)$
Hill model 4	4	$y = a \cdot \left(1 - \frac{(c-1) \cdot x^d}{b^d + x^d}\right)$
Inverse Exponential	4	$y = a \cdot (1 + (c-1)\exp(-bx^{-d}))$
Log-Normal Family	4	$y = a \cdot (1 + (c-1)\Phi(\ln b + d\ln x))$

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: relative.liver.weight..mean.

Fitted Models

model	converged	loglik	npar	AIC
full model	yes	191.91	13	-357.82
null model	yes	45.66	2	-87.32
Expon. m3-	yes	157.07	4	-306.14
Expon. m5-	yes	157.68	5	-305.36
Hill m3-	yes	157.08	4	-306.16
Hill m5-	yes	157.90	5	-305.80
Inv.Expon. m3-	yes	158.17	4	-308.34
Inv.Expon. m5-	yes	158.90	5	-307.80
LN m3-	yes	157.73	4	-307.46
LN m5-	yes	158.55	5	-307.10

Attention: the AIC of the best model (minimum AIC) is more than two units larger than that of the full model. This might indicate a problem in the data, in particular when the difference is much larger than two units (e.g. > 5). You might check the following options:1. In real-life studies, not all experimental factors are completely randomized over all animals (experimental units), e.g. animals were housed in the same cage within a given dose group, or order of treatments were not randomized over individual animals. Another option is that individual outlying animals distort the mean response of one or more treatment groups. This may lead to fluctuations in the (mean) responses among treatment groups that are larger than expected from random sampling error, resulting in an AIC difference with the full model larger than 2 units.2. the data consist of subgroups not taken into account in the model (e.g. various studies, or two sexes) 3. the data contain litter effects not taken into account 4. the response in the top dose group deviates substantially from the fitted model (check the CI around the observed (mean) response); Associated actions for each of these four options are:1. the greater scatter in (mean) responses will result in a wider BMD CI; normally, no further action is needed, as the BMD approach is relatively robust to such devations. You might check this by leaving out specific treatment groups (one by one) and check if this has a major impact on the BMD CI.2.

use the factor defining the subgroups as a covariate and re-analyse the data 3. re-analyse the data with litter effects taken into account 4. consider to leave out the top dose; it is not recommended to leave out two high dose groups.

Estimated Model Parameters

EXP

estimate for var- : 0.004272

estimate for a- : 34.59

estimate for CED- : 65.65

estimate for d- : 0.5886

HILL

estimate for var- : 0.004271

estimate for a- : 34.59

estimate for CED- : 65.65

estimate for d- : 0.5905

INVEXP

estimate for var- : 0.004194

estimate for a- : 34.67

estimate for CED- : 65.43

estimate for d- : 0.1165

LOGN

estimate for var- : 0.004225

estimate for a- : 34.63

estimate for CED- : 65.58

estimate for d- : 0.2091

Weights for Model Averaging

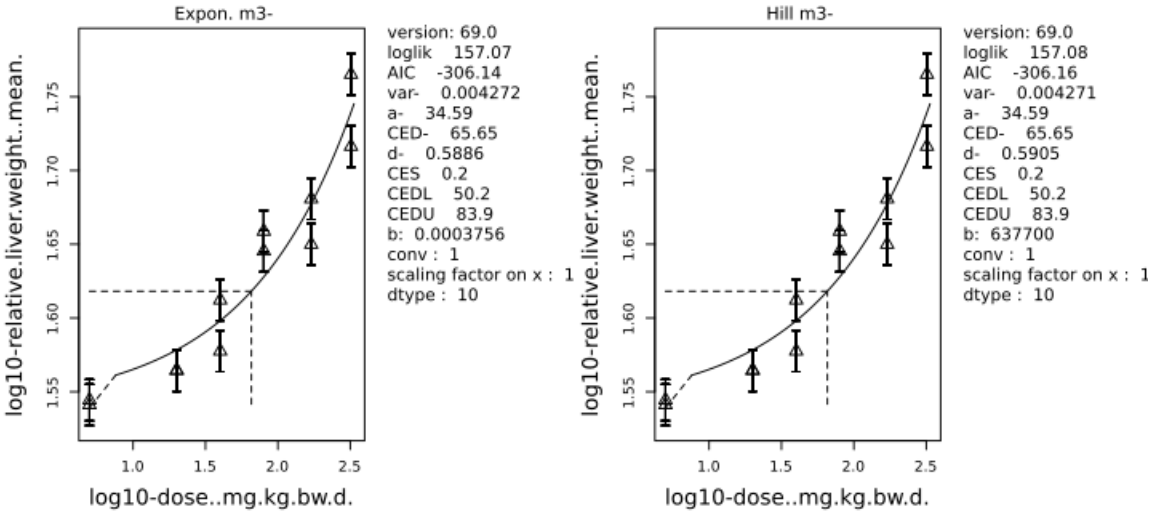
EXP	HILL	INVEXP	LOGN
0.14	0.15	0.43	0.28

Final BMD Values

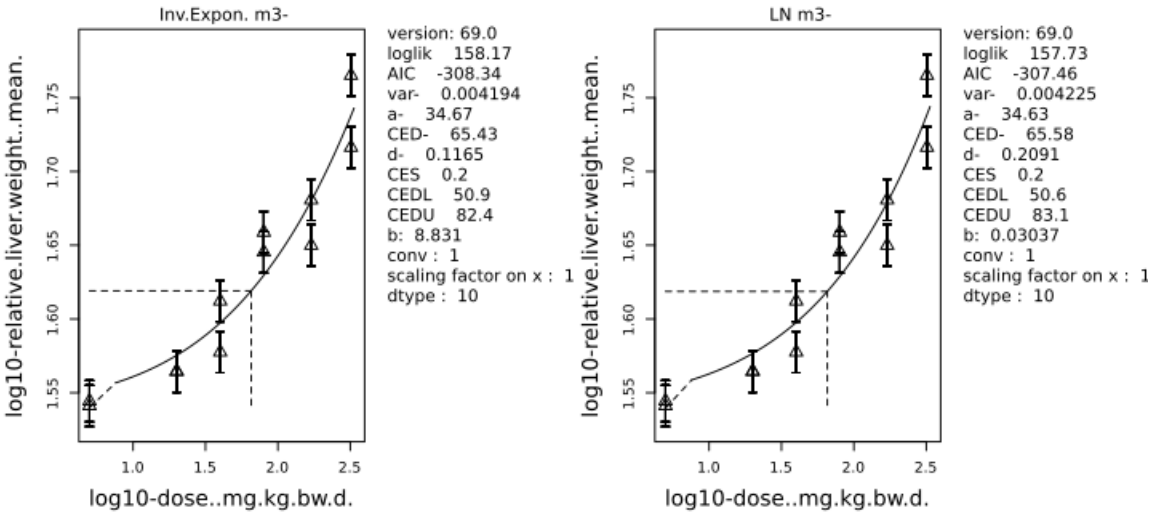
endpoint	subgroup	BMDL	BMDU
relative.liver.weight..mean.	all	49.3	83.2

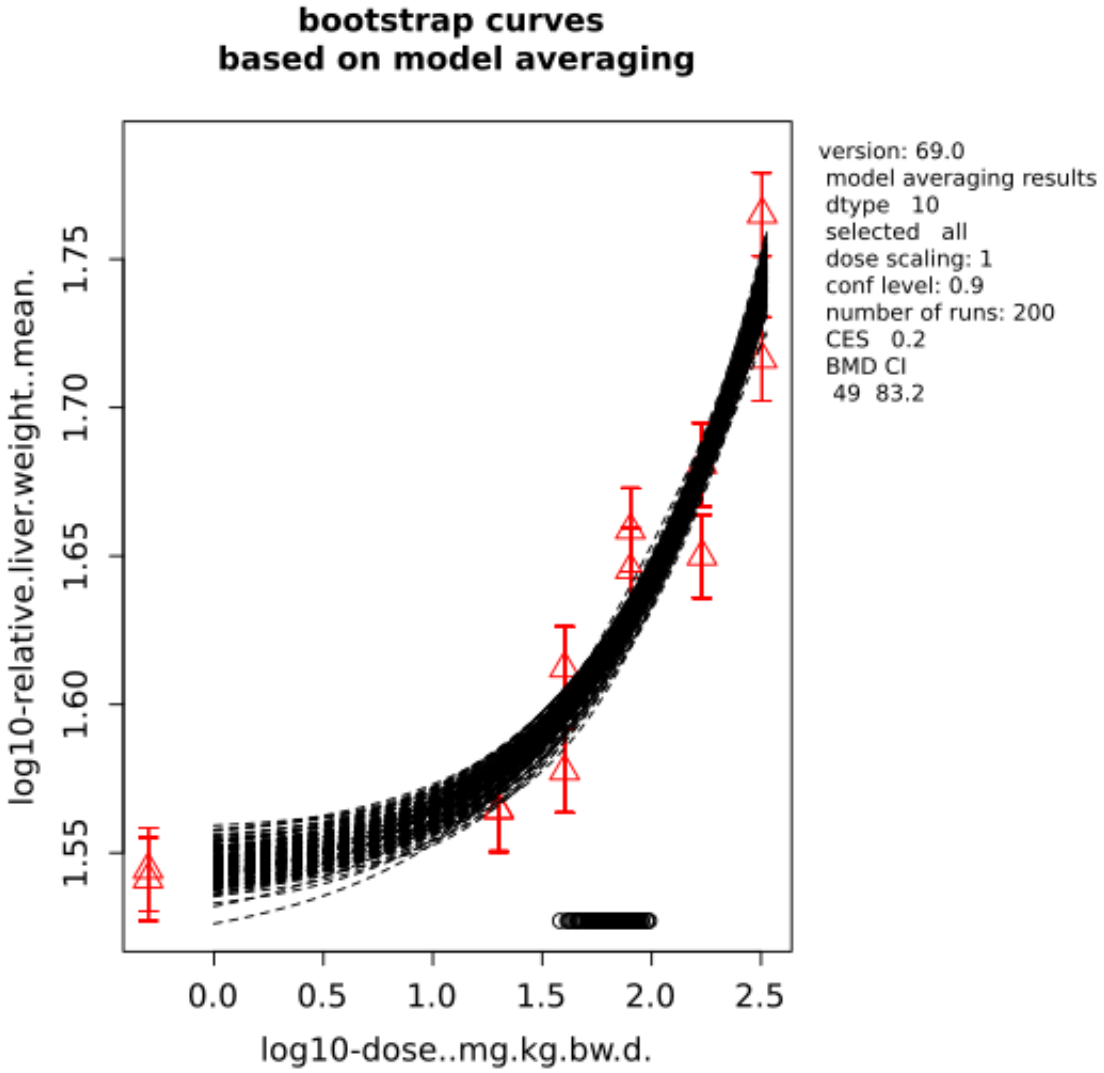
Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization



1,1,2,2 Tetrachloroethane





Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Appendix

References

Benchmark Dose Modeling: Report for N-octadecyl β-(3',5'-di-tert-butyl-4'-hydroxyphenyl) propionate (OBPP)

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: relative.liver.weight..mean..

Data used for analysis:

dose..mg.kg.bw.d.	relative.liver.weight..mean.	relative.liver.weight..standard.error.	n
0	5.0	0.1	6
30	5.6	0.1	5
100	6.0	0.2	5
300	6.9	0.1	5
1000	7.9	0.3	5

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is a 10% change in mean response compared to the controls. The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

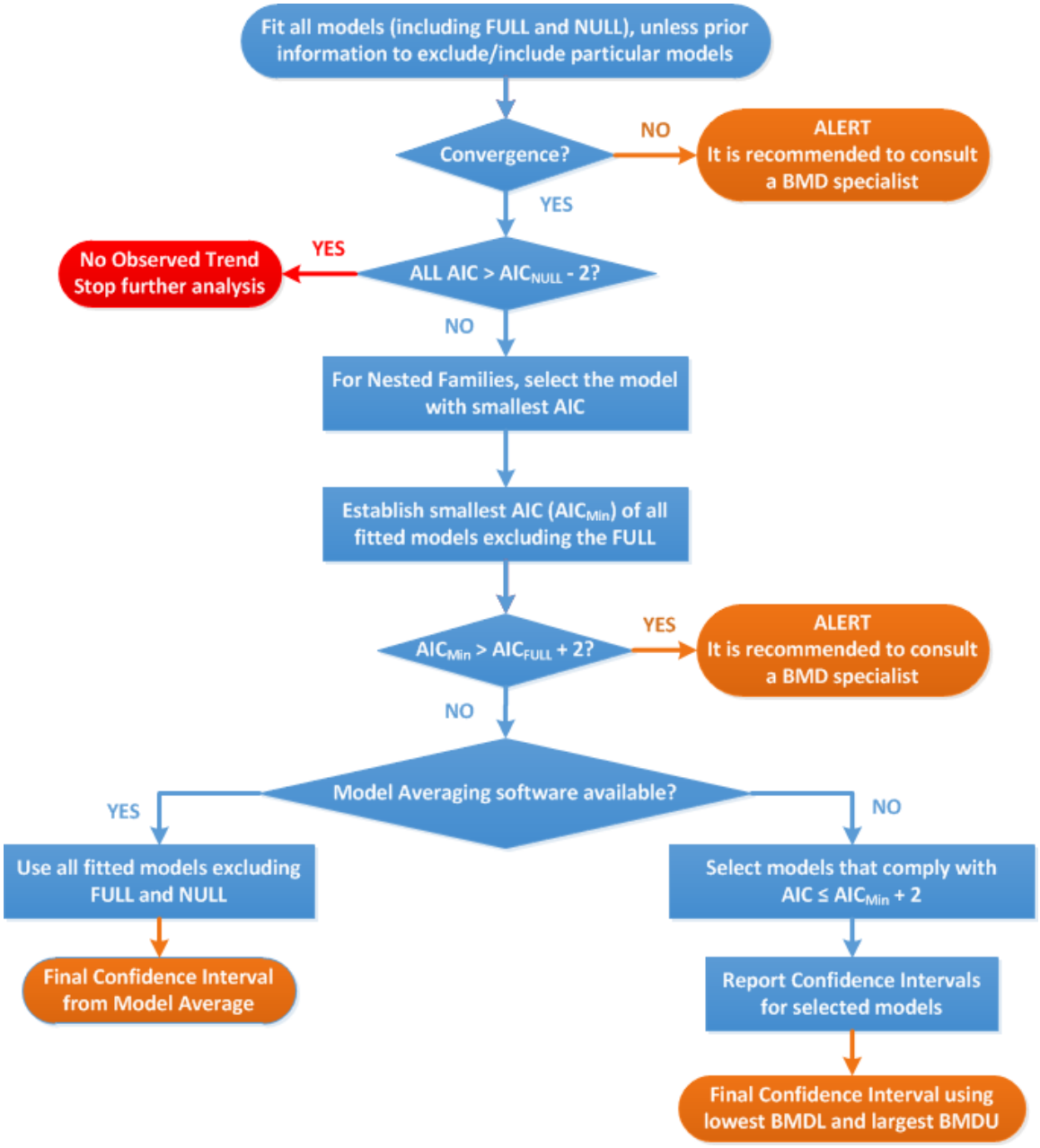
Default set of fitted models:

Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean
Exp model 3	3	$y = a \cdot \exp(bx^d)$
Exp model 4	4	$y = a \cdot (c - (c-1)\exp(-bx^d))$

Hill model 3	3	$y = a \cdot \left(1 - \frac{x^d}{b^d + x^d}\right)$
Hill model 4	4	$y = a \cdot \left(1 - \frac{(c-1) \cdot x^d}{b^d + x^d}\right)$
Inverse Exponential	4	$y = a \cdot (1 + (c-1)\exp(-bx^{-d}))$
Log-Normal Family	4	$y = a \cdot (1 + (c-1)\Phi(\ln b + d\ln x))$

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: relative.liver.weight..mean.

Fitted Models

model	converged	loglik	npar	AIC
full model	yes	39.41	6	-66.82
null model	yes	9.17	2	-14.34
Expon. m3-	yes	38.44	4	-68.88
Expon. m5-	yes	39.11	5	-68.22
Hill m3-	yes	38.44	4	-68.88
Hill m5-	yes	39.06	5	-68.12
Inv.Expon. m3-	yes	38.74	4	-69.48
Inv.Expon. m5-	yes	38.96	5	-67.92
LN m3-	yes	38.62	4	-69.24
LN m5-	yes	38.99	5	-67.98

Estimated Model Parameters

EXP

estimate for var- : 0.003044

estimate for a- : 4.977

estimate for CED- : 14.32

estimate for d- : 0.3759

HILL

estimate for var- : 0.003043

estimate for a- : 4.977

estimate for CED- : 14.38

estimate for d- : 0.3773

INVEXP

estimate for var- : 0.002975

estimate for a- : 4.985

estimate for CED- : 18.81

estimate for d- : 0.07606

LOGN

estimate for var- : 0.003001

estimate for a- : 4.982

estimate for CED- : 16.87

estimate for d- : 0.1352

Weights for Model Averaging

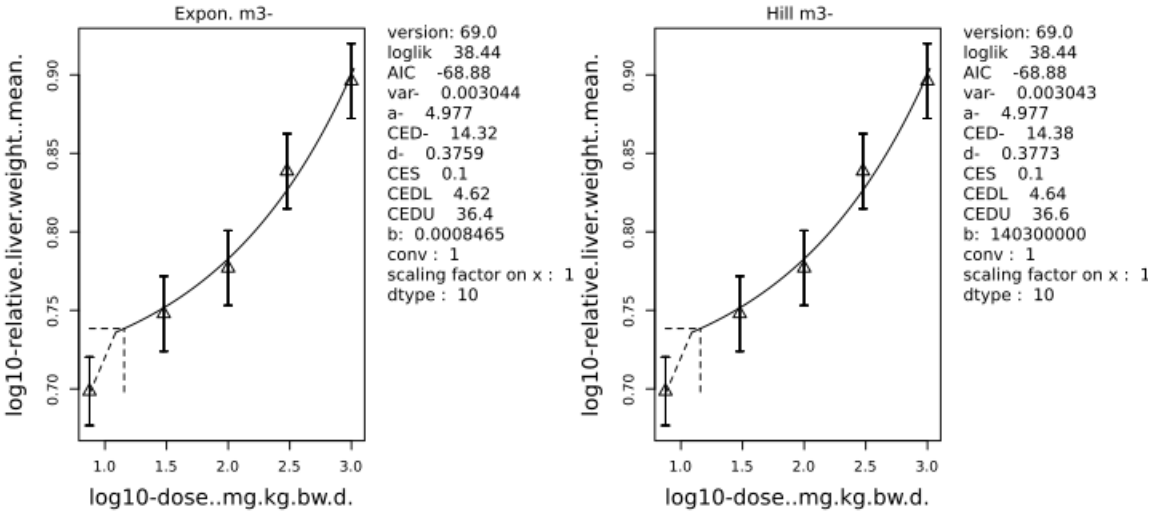
EXP	HILL	INVEXP	LOGN
0.22	0.22	0.3	0.26

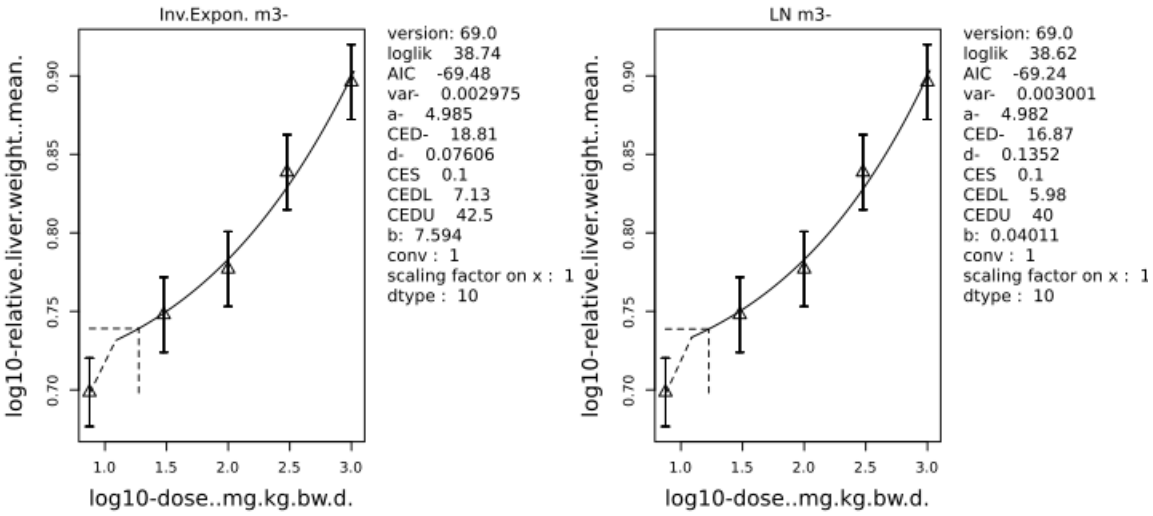
Final BMD Values

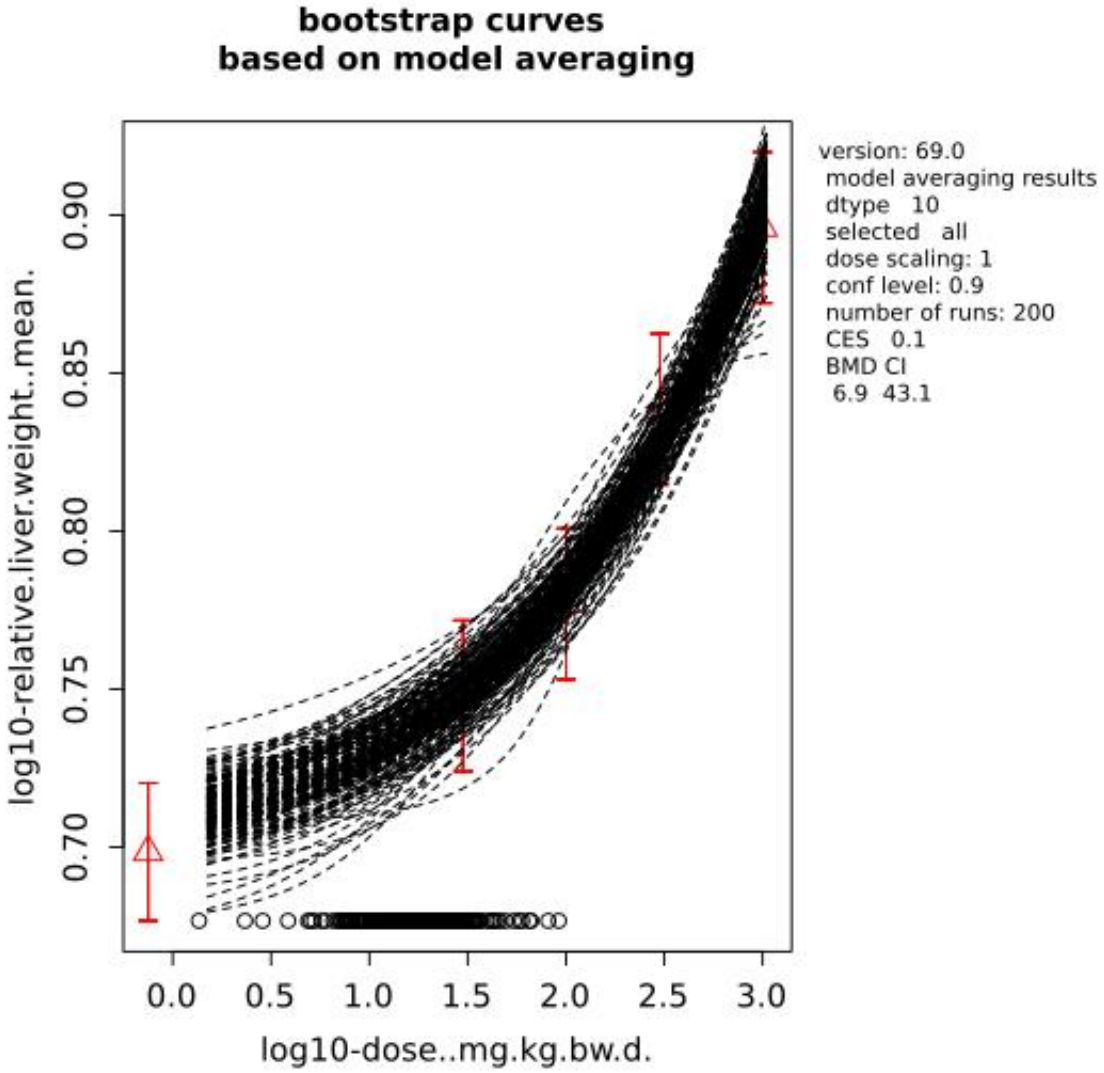
endpoint	subgroup	BMDL	BMDU
relative.liver.weight..mean.	all	6.86	43.1

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization




Inv.Expon. m3-
log10-relative.liver.weight..mean.
log10-dose..mg.kg.bw.d.
version: 69.0
loglik 38.74
AIC -69.48
var- 0.002975
a- 4.985
CED- 18.81
d- 0.07606
CES 0.1
CEDL 7.13
CEDU 42.5
b: 7.594
conv : 1
scaling factor on x : 1
dtype : 10
LN m3-
log10-relative.liver.weight..mean.
log10-dose..mg.kg.bw.d.
version: 69.0
loglik 38.62
AIC -69.24
var- 0.003001
a- 4.982
CED- 16.87
d- 0.1352
CES 0.1
CEDL 5.98
CEDU 40
b: 0.04011
conv : 1
scaling factor on x : 1
dtype : 10



Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: relative.liver.weight..mean..

Data used for analysis:

dose..mg.kg.bw.d.	relative.liver.weight..mean.	relative.liver.weight..standard.error.	n
0	5.0	0.1	6
30	5.6	0.1	5
100	6.0	0.2	5
300	6.9	0.1	5
1000	7.9	0.3	5

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is a 20% change in mean response compared to the controls. The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

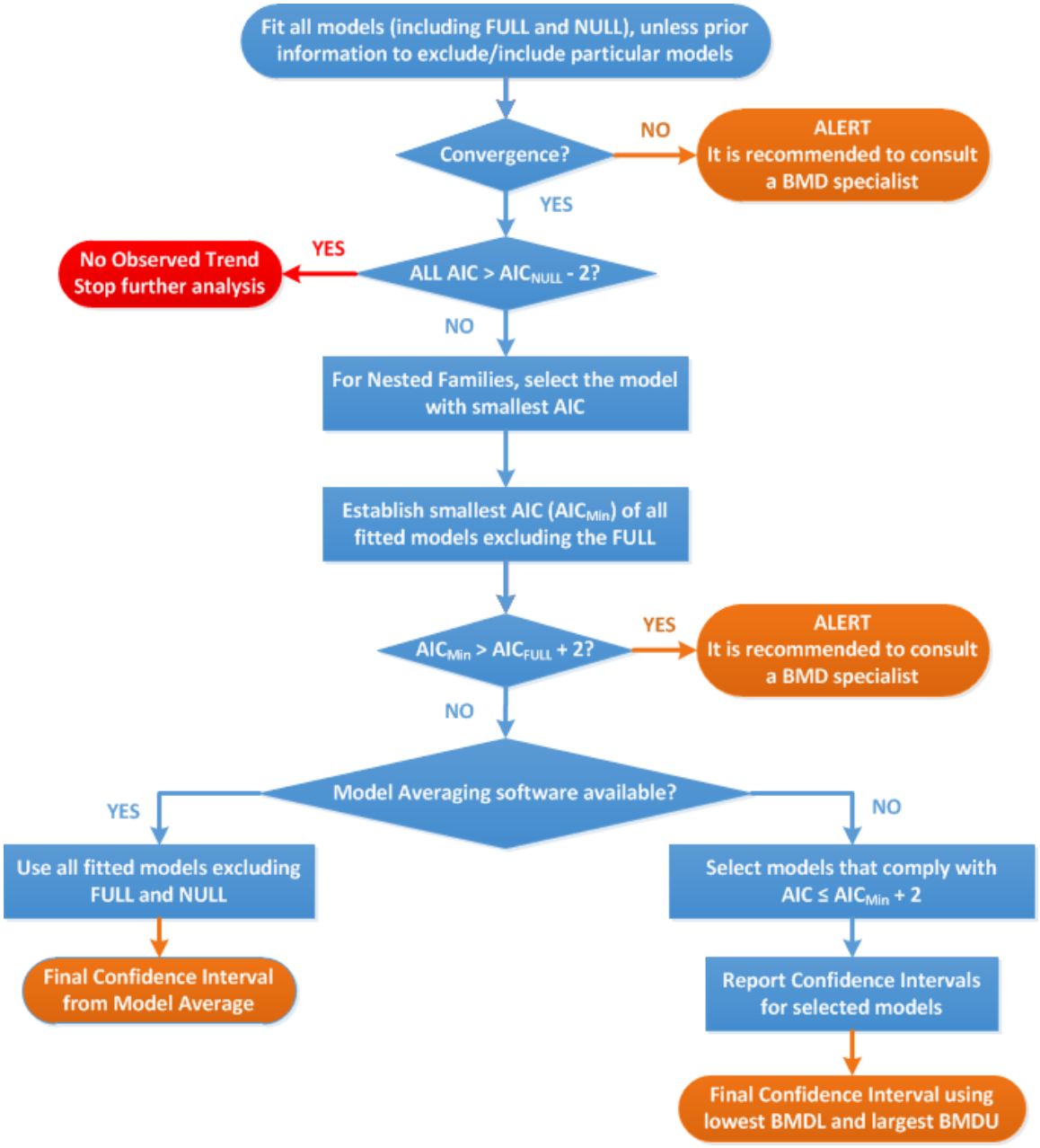
Default set of fitted models:

Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean
Exp model 3	3	$y = a \cdot \exp(bx^d)$
Exp model 4	4	$y = a \cdot (c - (c-1)\exp(-bx^d))$

Hill model 3	3	$y = a \cdot \left(1 - \frac{x^d}{b^d + x^d}\right)$
Hill model 4	4	$y = a \cdot \left(1 - \frac{(c-1) \cdot x^d}{b^d + x^d}\right)$
Inverse Exponential	4	$y = a \cdot (1 + (c-1)\exp(-bx^{-d}))$
Log-Normal Family	4	$y = a \cdot (1 + (c-1)\Phi(\ln b + d\ln x))$

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: relative.liver.weight..mean.

Fitted Models

model	converged	loglik	npar	AIC
full model	yes	39.41	6	-66.82
null model	yes	9.17	2	-14.34
Expon. m3-	yes	38.44	4	-68.88
Expon. m5-	yes	39.11	5	-68.22
Hill m3-	yes	38.44	4	-68.88
Hill m5-	yes	39.06	5	-68.12
Inv.Expon. m3-	yes	38.74	4	-69.48
Inv.Expon. m5-	yes	38.96	5	-67.92
LN m3-	yes	38.62	4	-69.24
LN m5-	yes	38.99	5	-67.98

Estimated Model Parameters

EXP

estimate for var- : 0.003044

estimate for a- : 4.977

estimate for CED- : 80.64

estimate for d- : 0.3759

HILL

estimate for var- : 0.003043

estimate for a- : 4.977

estimate for CED- : 80.69

estimate for d- : 0.3773

INVEXP

estimate for var- : 0.002975

estimate for a- : 4.985

estimate for CED- : 83

estimate for d- : 0.07606

LOGN

estimate for var- : 0.003001

estimate for a- : 4.982

estimate for CED- : 82.13

estimate for d- : 0.1352

Weights for Model Averaging

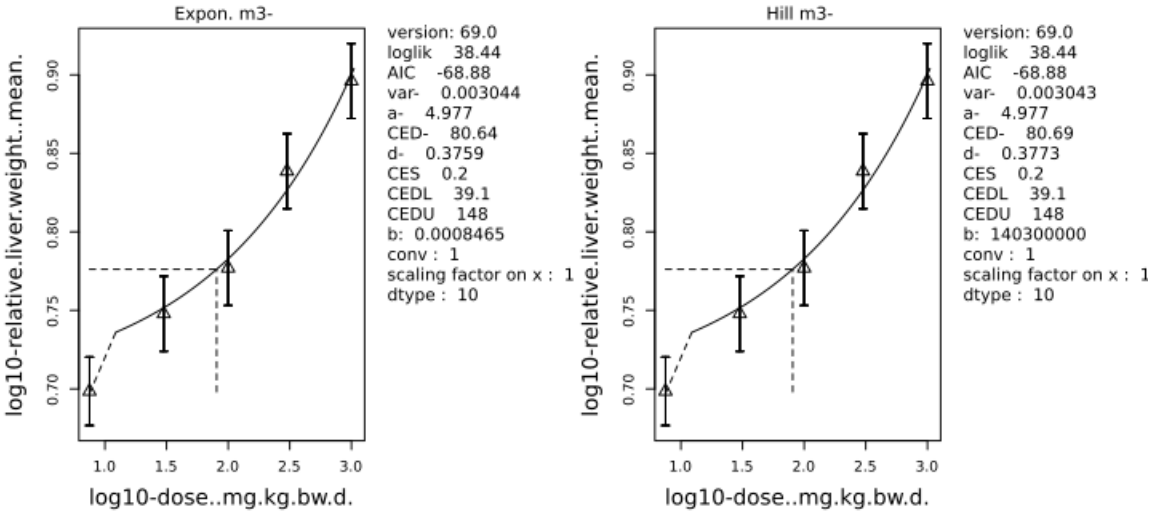
EXP	HILL	INVEXP	LOGN
0.22	0.22	0.3	0.26

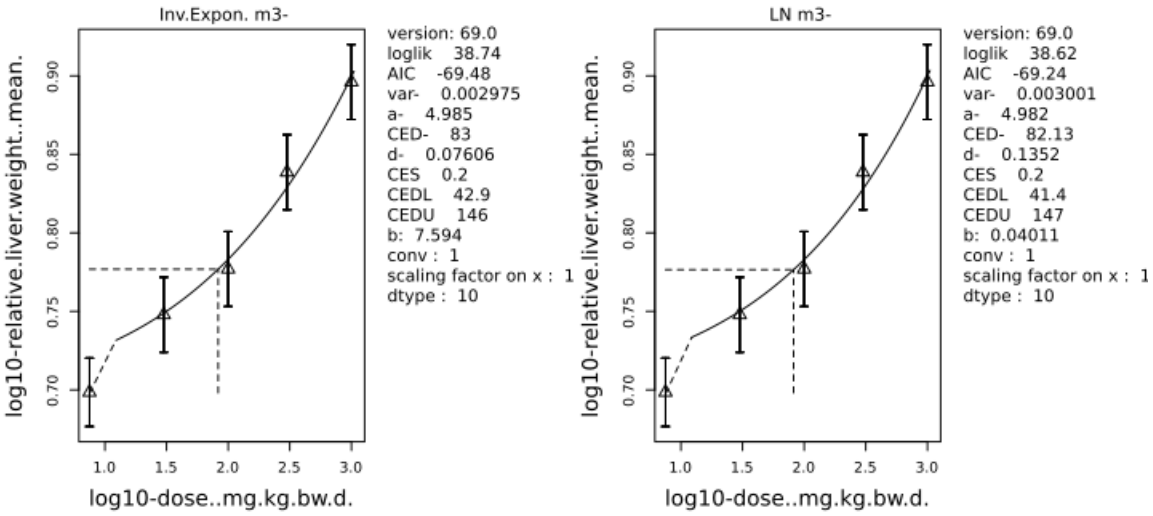
Final BMD Values

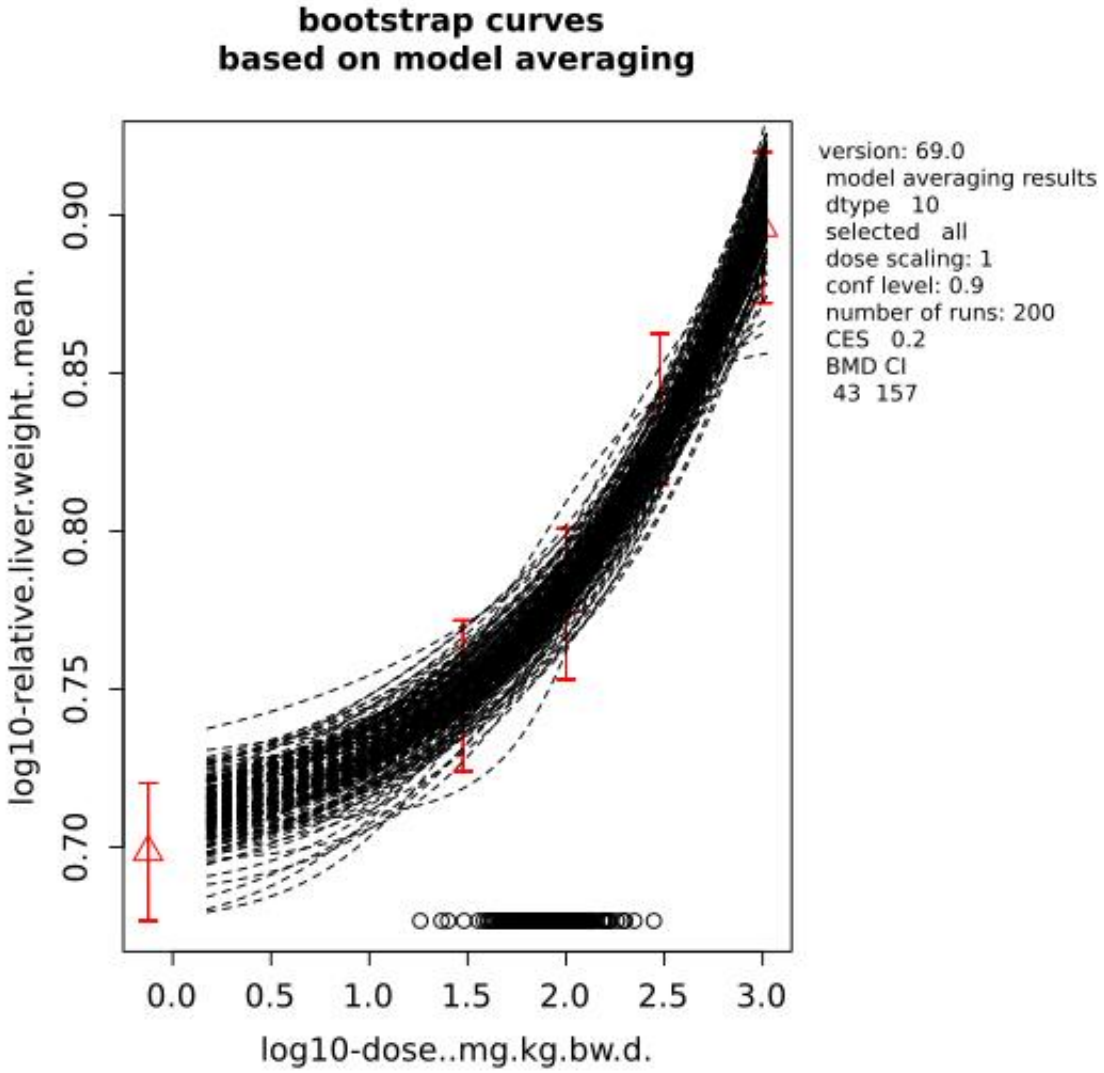
endpoint	subgroup	BMDL	BMDU
relative.liver.weight..mean.	all	42.6	157

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization




Inv.Expon. m3-
log10-relative.liver.weight..mean.
log10-dose..mg.kg.bw.d.
version: 69.0
loglik 38.74
AIC -69.48
var- 0.002975
a- 4.985
CED- 83
d- 0.07606
CES 0.2
CEDL 42.9
CEDU 146
b: 7.594
conv : 1
scaling factor on x : 1
dtype : 10
LN m3-
log10-relative.liver.weight..mean.
log10-dose..mg.kg.bw.d.
version: 69.0
loglik 38.62
AIC -69.24
var- 0.003001
a- 4.982
CED- 82.13
d- 0.1352
CES 0.2
CEDL 41.4
CEDU 147
b: 0.04011
conv : 1
scaling factor on x : 1
dtype : 10



Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for Tert-Butyl alcohol

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: relative.kidney.weight..mean..

Data used for analysis:

dose..mg.kg.bw.d.	relative.kidney.weight..mean.	relative.kidney.weight..standard.error.	n
0	3.49	0.08	10
180	3.99	0.07	10
330	4.21	0.08	10
650	4.95	0.17	10

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is a 7% change in mean response compared to the controls. The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 67.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

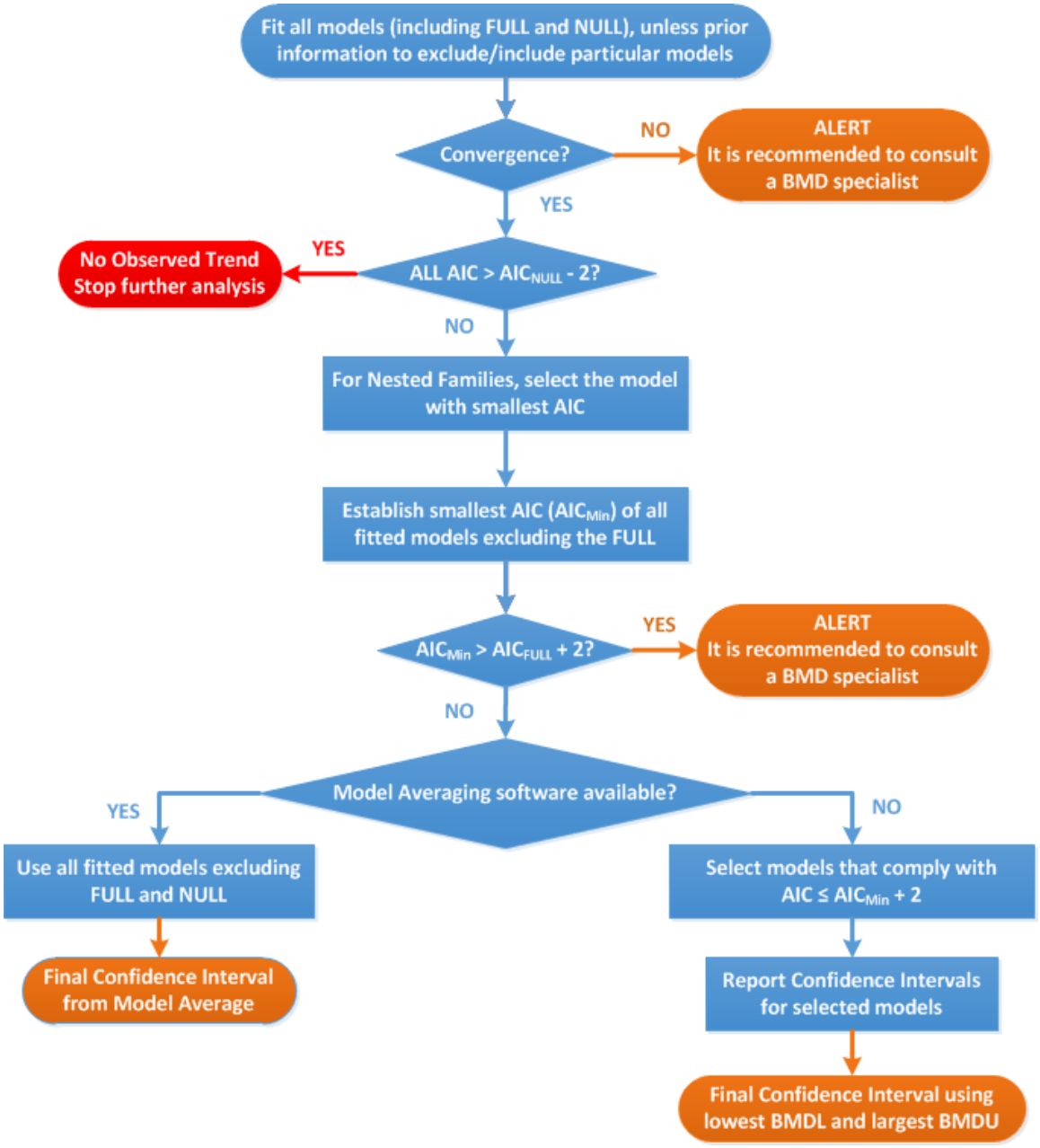
Default set of fitted models:

Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean
Exp model 3	3	$y = a \cdot \exp(bx^d)$
Exp model 4	4	$y = a \cdot (c - (c-1)\exp(-bx^d))$

Hill model 3	3	$y = a \cdot \left(1 - \frac{x^d}{b^d + x^d}\right)$
Hill model 4	4	$y = a \cdot \left(1 - \frac{(c-1) \cdot x^d}{b^d + x^d}\right)$
Inverse Exponential	4	$y = a \cdot (1 + (c-1)\exp(-bx^{-d}))$
Log-Normal Family	4	$y = a \cdot (1 + (c-1)\Phi(\ln b + d\ln x))$

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: relative.kidney.weight..mean.

Fitted Models

model	converged	loglik	npar	AIC
full model	yes	47.97	5	-85.94
null model	yes	20.80	2	-37.60
Expon. m3-	yes	47.71	4	-87.42
Expon. m5-	yes	47.71	5	-85.42
Hill m3-	yes	47.71	4	-87.42
Hill m5-	yes	47.70	5	-85.40
Inv.Expon. m3-	yes	47.59	4	-87.18
Inv.Expon. m5-	yes	47.55	5	-85.10
LN m3-	yes	47.64	4	-87.28
LN m5-	yes	47.62	5	-85.24

Estimated Model Parameters

EXP

estimate for var- : 0.005389

estimate for a- : 3.486

estimate for CED- : 85.13

estimate for d- : 0.7976

HILL

estimate for var- : 0.005389

estimate for a- : 3.486

estimate for CED- : 85.26

estimate for d- : 0.7998

INVEXP

estimate for var- : 0.005423

estimate for a- : 3.487

estimate for CED- : 96.34

estimate for d- : 0.1517

LOGN

estimate for var- : 0.005407

estimate for a- : 3.487

estimate for CED- : 91.65

estimate for d- : 0.2768

Weights for Model Averaging

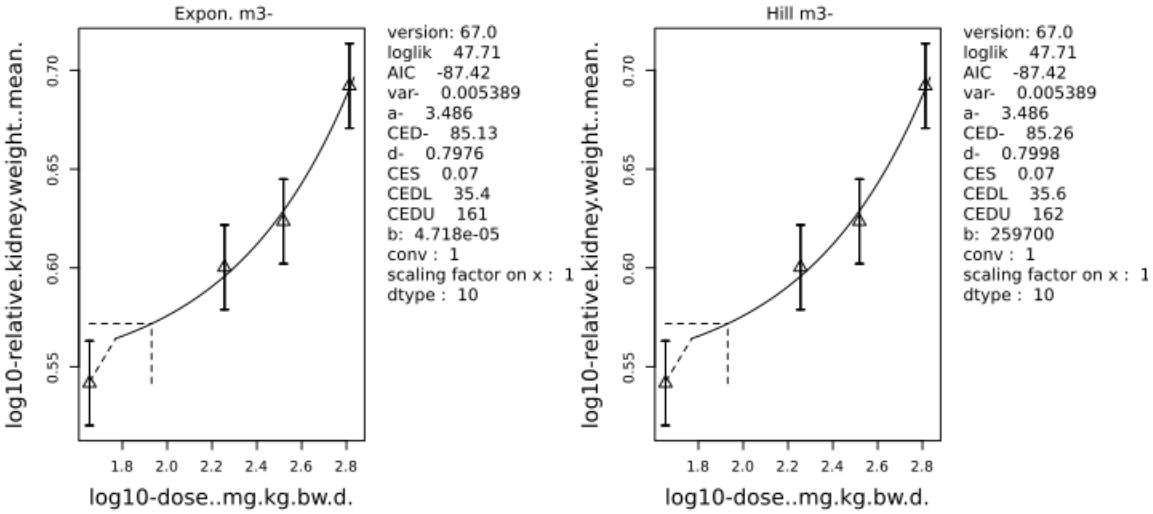
EXP	HILL	INVEXP	LOGN
0.26	0.26	0.23	0.24

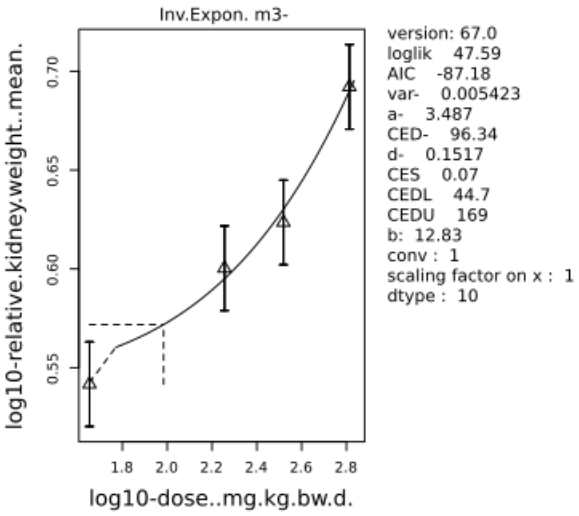
Final BMD Values

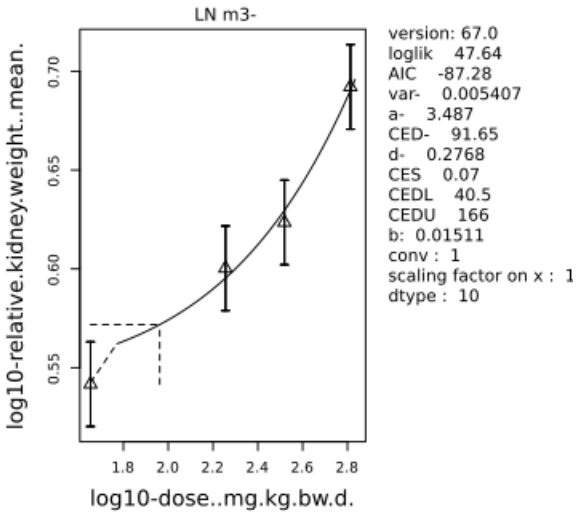
endpoint	subgroup	BMDL	BMDU
relative.kidney.weight..mean.		47.7	181

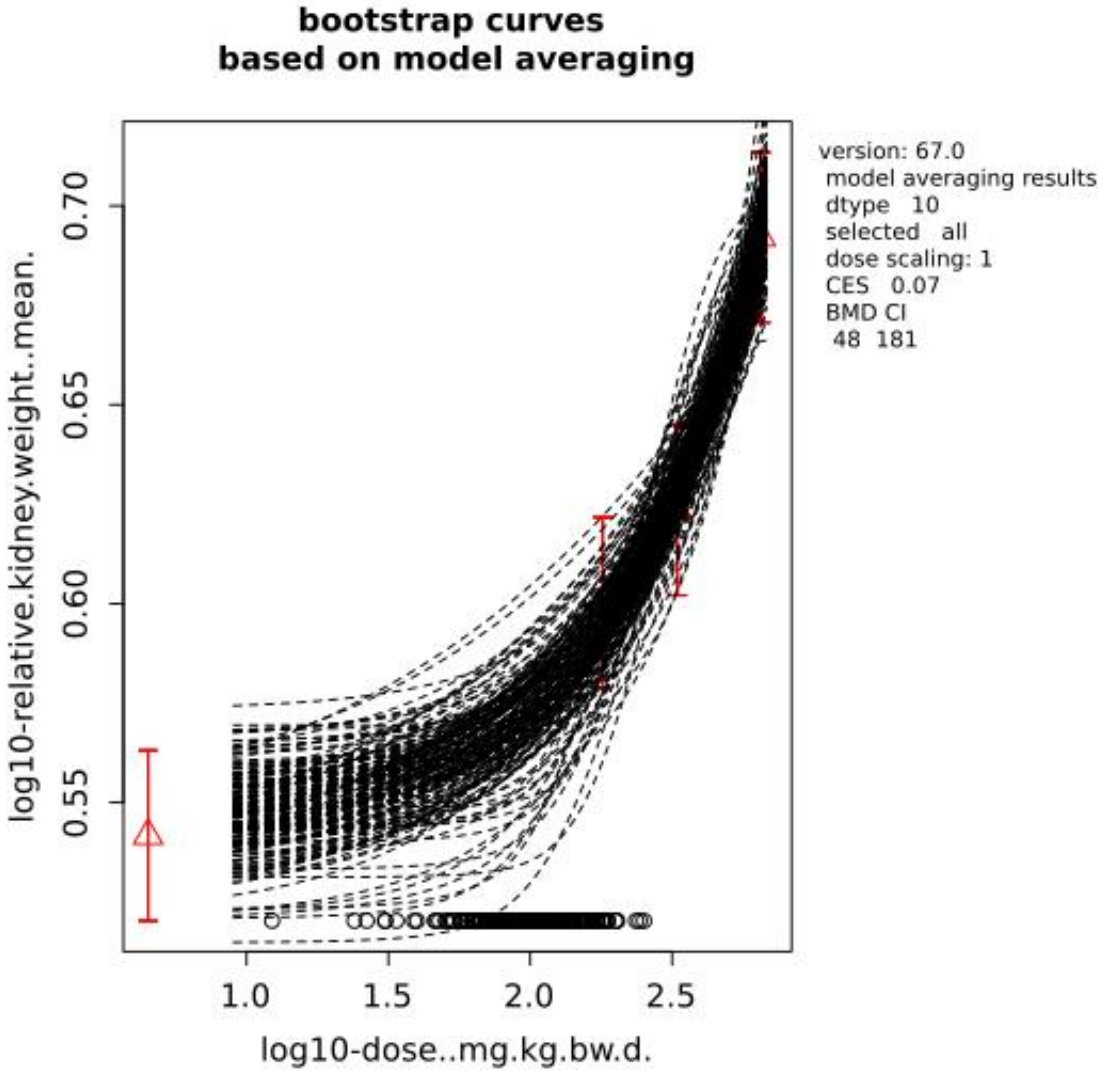
Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization

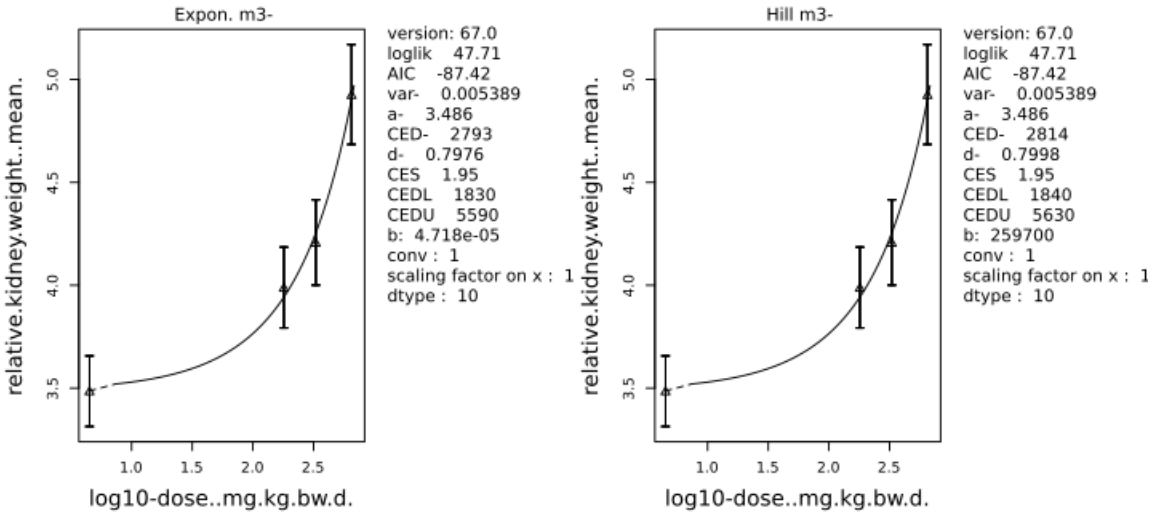



Inv.Expon. m3-
log10-relative.kidney.weight..mean.
0.55
0.60
0.65
0.70
1.8
2.0
2.2
2.4
2.6
2.8
log10-dose..mg.kg.bw.d.
version: 67.0
loglik 47.59
AIC -87.18
var- 0.005423
a- 3.487
CED- 96.34
d- 0.1517
CES 0.07
CEDL 44.7
CEDU 169
b: 12.83
conv : 1
scaling factor on x : 1
dtype : 10

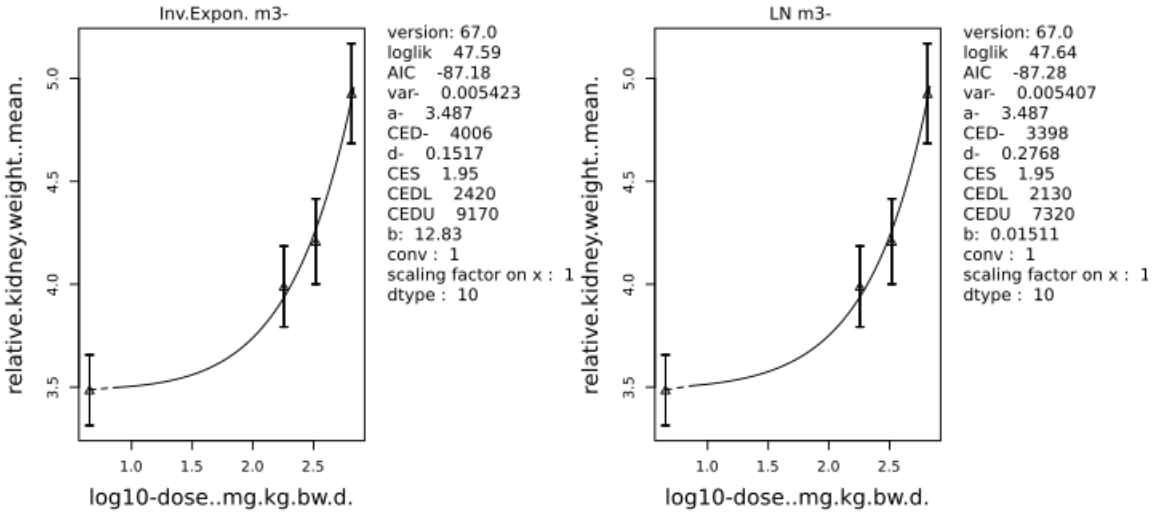

LN m3-
log10-relative.kidney.weight..mean.
0.55
0.60
0.65
0.70
1.8
2.0
2.2
2.4
2.6
2.8
log10-dose..mg.kg.bw.d.
version: 67.0
loglik 47.64
AIC -87.28
var- 0.005407
a- 3.487
CED- 91.65
d- 0.2768
CES 0.07
CEDL 40.5
CEDU 166
b: 0.01511
conv : 1
scaling factor on x : 1
dtype : 10



Advanced Plots



Advanced plots for response 'relative.kidney.weight..mean.'



Advanced plots for response 'relative.kidney.weight..mean.'

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

Benchmark Dose Modeling: Report for Benzene

European Food Safety Authority (EFSA)

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Key words: (max. seven key words)

Requestor: add requesting party

Question number: EFSA-Q-YYYY-NNNNN

Correspondence: xxx@efsa.europa.eu

Acknowledgements: [Scientific Committee OR EFSA] wishes to thank the following for the support provided to this scientific output: [staff members or others who made a contribution but are not eligible as authors]. The Panel [Scientific Committee OR EFSA] wishes to acknowledge all European competent institutions, Member State bodies and other organisations that provided data for this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Individual authors [add names in the format Surname followed by Initial(s), Surname followed by Initial(s) and Surname followed by Initial(s)], 20YY. Title of the report. EFSA supporting publication 20YY:EN-NNNN. 10 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

ISSN: 2397-8325

Summary

The summary should not include tables, footnotes, graphs or pictures or references.

Table of Contents

Data Description

The endpoint to be analyzed is: Abnormality.

Data used for analysis:

dose..ppm.y.	Abnormality	n
0.00	4	94
3.55	7	65
6.51	6	65
10.72	14	65
20.02	16	65
40.71	16	65

Information pertaining to this endpoint.

Selection of the BMR

The BMR (benchmark response) used is an extra risk of 10% compared to the controls.

When the specified BMR deviates from the default value, the rationale behind the choice made should be described.

The BMD (benchmark dose) is the dose corresponding with the BMR of interest.

A 90% confidence interval around the BMD will be estimated, the lower bound is reported by BMDL and the upper bound by BMDU.

Software Used

Results are obtained using the EFSA web-tool for BMD analysis, which uses the R-package PROAST, version 69.0, for the underlying calculations.

Specification of Deviations from Default Assumptions

General assumptions

Please motivate in detail assumptions made when deviating from the recommended defaults (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).

Dose-response models

Other models than the recommended ones that were fitted should be listed, with the respective description of reasons to include them.

Default set of fitted models:

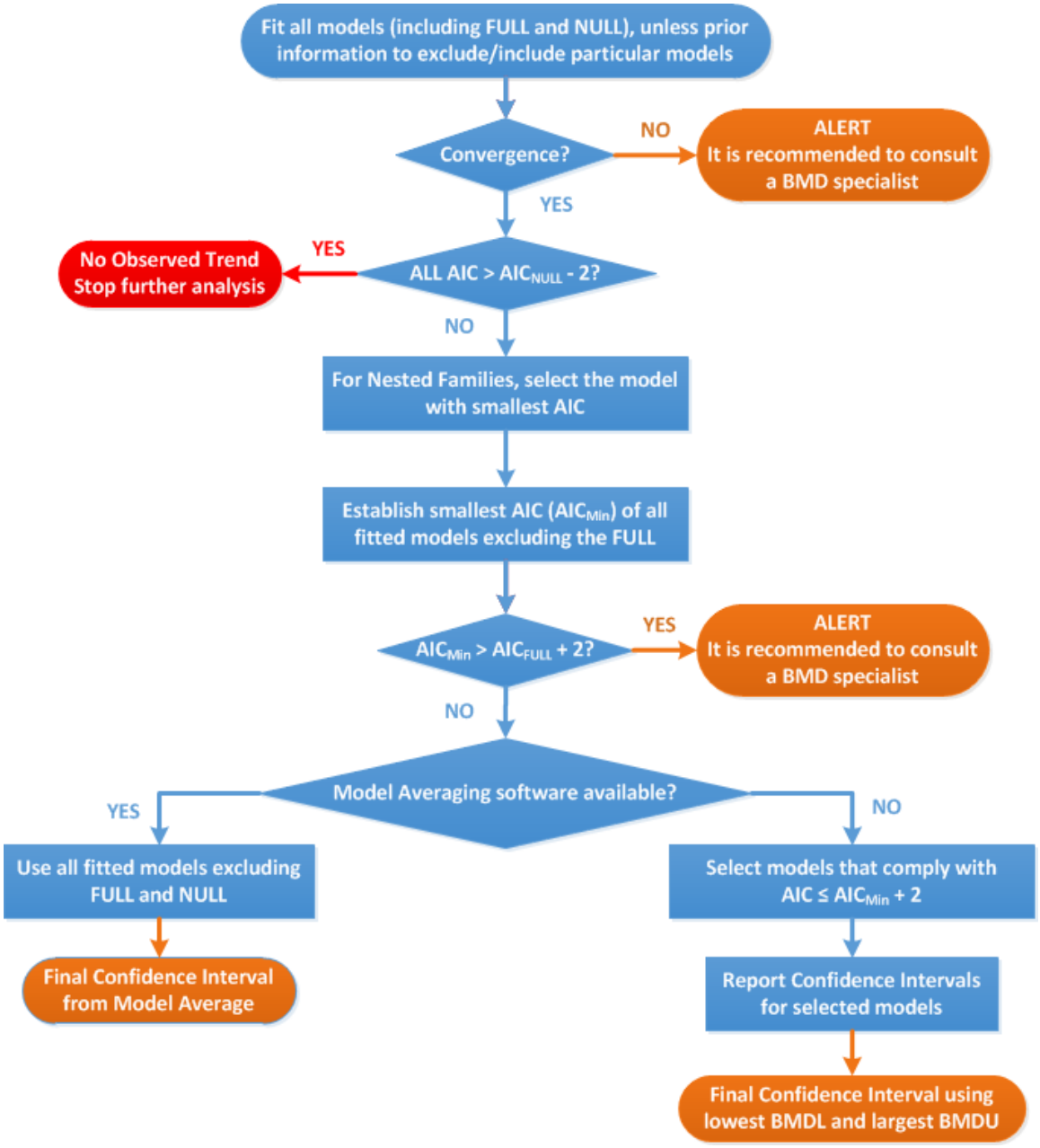
Model	Number of parameters	Formula
Null	1	$y = a$
Full	no. of groups	$y =$ group mean

Logistic	2	$y = \frac{1}{1 + \exp(-a - bx)}$
Probit	2	$y = pnorm((x - a) \cdot b)$
Log-logistic	3	$y = a + \frac{1 - a}{1 + \exp\left(c \cdot \log\left(\frac{b}{x}\right)\right)}$
Log-probit	3	$y = a + (1 - a) \cdot pnorm\left(c \cdot \log\left(\frac{x}{b}\right)\right)$
Weibull	3	$y = a + (1 - a)\left(1 - \exp\left(-\left(\frac{x}{b}\right)^{c}\right)\right)$
Gamma	3	$y = pgamma(bx; c)$
Two-stage	3	$y = a + (1 - a)\left(1 - \exp\left(-\frac{x}{b} - c\left(\frac{x}{b}\right)^{2}\right)\right)$
Exp model 3	3	$y = a \cdot \exp(bx^{d})$
Exp model 5	4	$y = a \cdot (c - (c - 1)\exp(-bx^{d}))$
Hill model 3	3	$y = a \cdot \left(1 - \frac{x^{d}}{b^{d} + x^{d}}\right)$
Hill model 5	4	$y = a \cdot \left(1 + (c - 1)\frac{x^{d}}{b^{d} + x^{d}}\right)$

For the Exp and Hill family, we fit models with 3 and 4 parameters as listed in the table. The 3-parameter model is selected if the difference in AIC is smaller than 5, otherwise the 4-parameter model is selected.

Procedure for selection of BMDL

Description of any deviation from the procedure described in the flow chart to obtain the final BMD confidence interval.



Flowchart for selection of BMDL

Results

Response variable: Abnormality

Fitted Models

model	No.par	loglik	AIC	accepted	BMDL	BMDU	BMD	conv
null	1	-177.38	356.76		NA	NA	NA	NA
full	6	-165.17	342.34		NA	NA	NA	NA
two.stage	3	-168.45	342.90	no	NA	NA	13.40	yes
log.logist	3	-166.72	339.44	yes	1.180	14.9	6.26	yes
Weibull	3	-166.77	339.54	yes	1.070	15.1	6.18	yes
log.prob	3	-166.65	339.30	yes	1.380	14.6	6.38	yes
gamma	3	-166.80	339.60	yes	0.984	15.3	6.16	yes
logistic	2	-170.49	344.98	no	NA	NA	22.50	yes
probit	2	-170.23	344.46	no	NA	NA	21.30	yes
LVM: Expon. m3-	3	-166.87	339.74	yes	2.720	15.6	5.92	yes
LVM: Hill m3-	3	-166.84	339.68	yes	2.270	15.4	5.99	yes

Estimated Model Parameters

two.stage

estimate for a- : 0.06582

estimate for BMD- : 13.43

estimate for c : 1e-06

log.logist

estimate for a- : 0.04117

estimate for BMD- : 6.255

estimate for c : 0.5787

Weibull

estimate for a- : 0.04113

estimate for BMD- : 6.182

estimate for c : 0.525

log.prob

estimate for a- : 0.04148

estimate for BMD- : 6.384

estimate for c : 0.3225

gamma

estimate for a- : 0.04118

estimate for BMD- : 6.161

estimate for c : 0.4852

logistic

estimate for a- : -2.243

estimate for BMD- : 22.48

probit

estimate for a- : -1.325

estimate for BMD- : 21.26

EXP

estimate for a- : 1.546

estimate for BMD- : 5.922

estimate for d- : 0.2988

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

HILL

estimate for a- : 1.546

estimate for BMD- : 5.995

estimate for d- : 0.3354

estimate for th(fixed) : 0

estimate for sigma(fixed) : 0.25

Weights for Model Averaging

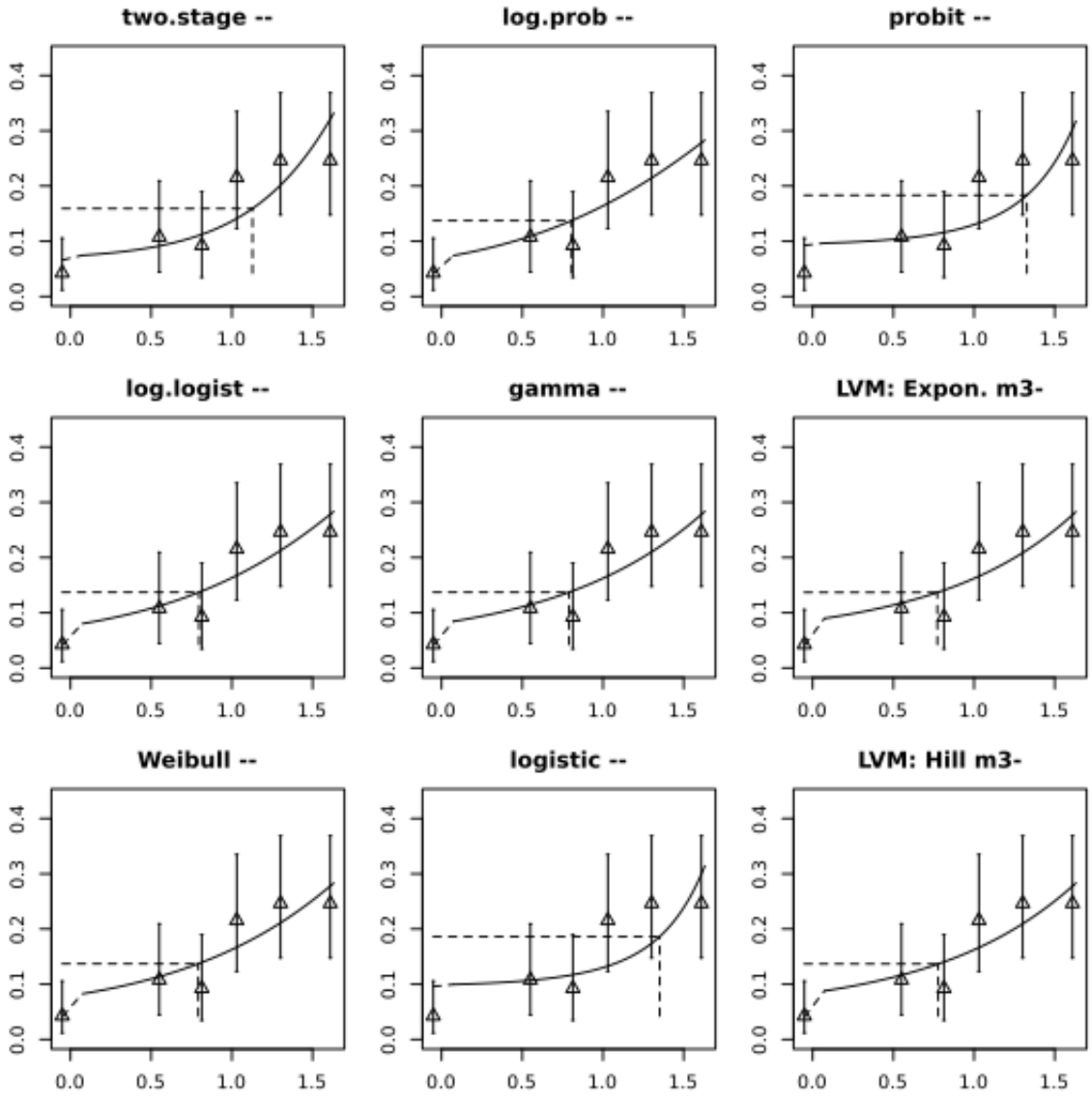
two.stage	log.logist	Weibull	log.prob	gamma	logistic	probit	EXP	HILL
0.03	0.17	0.16	0.18	0.15	0.01	0.01	0.14	0.15

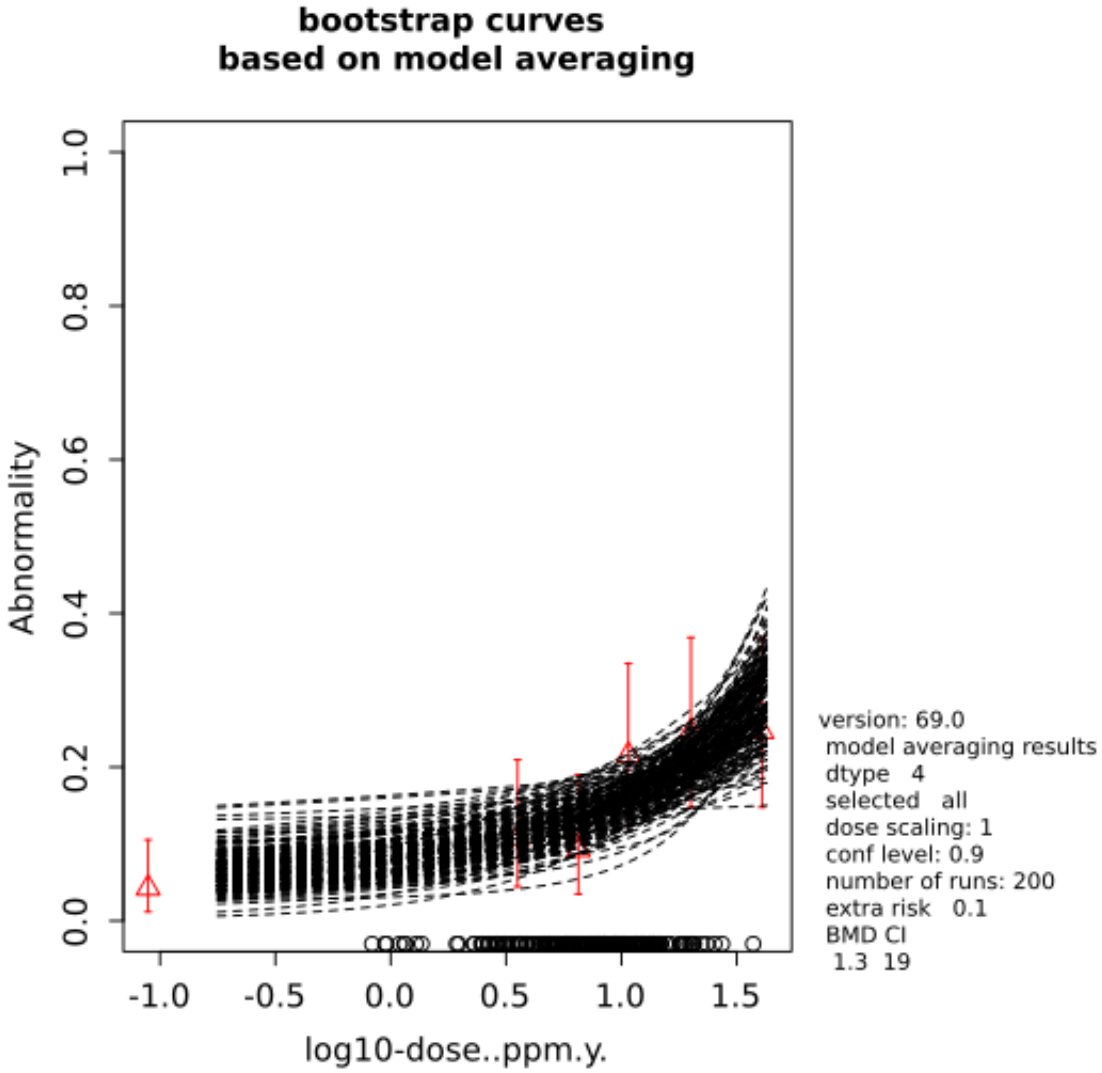
Final BMD Values

subgroup	BMDL	BMDU
all	1.28	19

Confidence intervals for the BMD are based on 200 bootstrap data sets.

Visualization





Advanced Plots

No results available: If needed, please create advanced plots in the application.

Conclusions

The section should discuss the results for the different endpoints and, if applicable, specific issues such as:

- *Discuss if there were any alerts, and if so, how they well dealt with.*
- *Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.*

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.

REPORT 4:
Probabilistic Hazard Assessment

RESEARCH PROJECT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Summary

In probabilistic approaches the input data to the following equation are given as distributions, expressing uncertainty (in the POD and the assessment factors) and variability (in the human population) of the input parameters:

$$GV = \frac{POD}{AF_1 * AF_2 * AF_3 + \cdots}$$

with: GV = guidance value, POD = point of departure, AF = assessment factor.

These distributions are combined using probabilistic methods (Monte Carlo analysis), resulting in a distribution of GV.

This approach requires to decide on the critical effect size or benchmark response (BMR), in order to determine the POD, if a benchmark dose is used, on the percentage of the target population to be covered by GV and on the probability of achieving the defined protection level.
The distribution of GV then allows describing uncertainty and variability of the output, to better characterise the protection level achieved and to estimate the size and likeliness of health effects at higher concentrations.

With regard to the practical implementation of probabilistic approaches in risk assessment two major recent developments are described in this report:

- The APROBA tool developed in the frame of the WHO/IPCS project on “Evaluating and Expressing Uncertainty in Hazard Characterization” (WHO 2014).

This EXCEL®-based tool allows rapid approximations to full Monte Carlo analyses by using lognormal input distributions for all parameters.

- The Monte Carlo tool developed by EFSA

This full Monte Carlo analysis tool is currently under development at EFSA and allows not only Monte Carlo analyses, but also distribution fitting and use of various kinds of distributions.

The principles, pros and cons of the various approaches, the input data used by APROBA and the implications of different types of dose-response data are discussed in this report. With two simple examples the workability of the tools are demonstrated and results are compared.

In view of these new developments use of probabilistic approaches to hazard assessment is simplified and their use for

- method development and discussion of (combination of) deterministic factors
- comparison with standard assessments using deterministic factors
- refined assessments of complex cases

is encouraged.

Abbreviations

AEL	Acceptable Exposure Levels
AAEL	Acute Acceptable Exposure Levels
AF	Assessment factor
AGS	Ausschuss für Gefahrstoffe
AIC	Akaike information criterion
ANSES	Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail
AOEL	Acceptable Operator Exposure Levels
AAOEL	Acute Acceptable Operator Exposure Levels
APROBA	Approximate probabilistic analysis
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
BBMD	Bayesian Benchmark Dose
BMD	Benchmark dose
BMDL	Benchmark dose lower bound
BMDU	Benchmark dose upper bound
BMR	Benchmark response
BMDS	Benchmark dose software
BOELV	Binding occupational exposure level values
BPR	Biocidal products regulation
BS	Bootstrapping
CDS	Cumulative distribution function
CES	Critical effect size
CSAF	Chemical-specific adjustment factors

DFG	Deutsche Forschungsgesellschaft
DMEL	Derived minimal effect level
DNEL	Derived no effect level
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
ED10	Effective dose 10% (dose corresponding to a 10% increase in an adverse effect, relative to the control response)
EFSA	European Food Safety Authority
GM	Geometric mean
GSD	Geometric standard deviation
GV	Guidance value
IPCS	WHO's International Programme on Chemical Safety
IRIS	Integrated Risk Information System
LOAEC	Lowest observed adverse effect concentration
LOAEL	Lowest observed adverse effect level
MAK	Maximale Arbeitsplatzkonzentration
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MCRA	Monte Carlo Risk Assessment
MPPD	Multiple path particle dosimetry (model)
NAEC	No adverse effect concentration
NAEL	No adverse effect level
NOAEC	No observed adverse effect concentration
NOAEL	No observed adverse effect level

OEL	Occupational exposure limit
PBPK	Physiology-based pharmacokinetic (model)
PDF	Probability density function
POD	Point of departure
PPP	Plant protection products
PROAST	Dose-response modelling software by RIVM
QSAR	Quantitative structure activity relationship
RAC	Committee for Risk Assessment
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals,
RfD	Reference dose
RIVM	Dutch National Institute for Public Health and the Environment
SC	EFSA's Scientific Committee
SCOEL	Scientific Committee on Occupational Exposure Limits
STEL	Short-term exposure limit
SD	Standard deviation
TD	Toxicodynamics
TK	Toxicokinetics
TRGS	Technische Regeln für Gefahrstoffe
US EPA	Environmental Protection Agency in the US
WHO	World Health Organisation

1 Introduction

Health-based guidance values such as OELs are typically derived by dividing a point of departure (POD) by assessment factors (including allometric scaling factors) accounting for uncertainty and variability. In addition, adjustments, e.g. for differences in exposure conditions, might be considered.

Uncertainty

Uncertainty in this context means uncertainty of any input data due to e.g. inaccurate measurements or unknown representability of entity of tested chemicals for the target substance. Uncertainty can be reduced by improving input data.

Variability

Variability means the variation of e.g. internal doses and responses in the target population. This is a real property of the target population and cannot be reduced.

In the so-called deterministic approach, a single definite value is used for each input variable. If input data are properly chosen, the resulting value is thought to provide sufficient protection against harmful effects of the evaluated chemical. But such point estimates have some significant disadvantages:

- They do not come with a specification of the level of protection achieved
- They do not inform (quantitatively) about remaining uncertainties
- If for the applied assessment factors conservative estimates are used, multiplication of these conservative factors might lead to an overly conservative guidance value; on the other hand, using e.g., average values for each assessment factor (e.g., a factor sufficient to cover 50% of chemicals in an empirical dataset), might lead to a guidance value not sufficiently protective in the majority of cases.

Since several years so-called probabilistic approaches for deriving health-based guidance values are proposed (Mekel and Fehr 2014). These approaches are characterised as follows:

- Input variables are not deterministic values but distributions (e.g., the distribution of possible values for extrapolating from subchronic to chronic study duration; these distributions are typically obtained from empirical datasets)
- These input distributions describe variability and/or uncertainty in the input data

As no discrete mathematical solutions are available for combining distributions of various forms, they need to be combined using probabilistic methods (Monte-Carlo analysis) (Aral 2010). This approach is graphically presented in **Figure 1-1**.

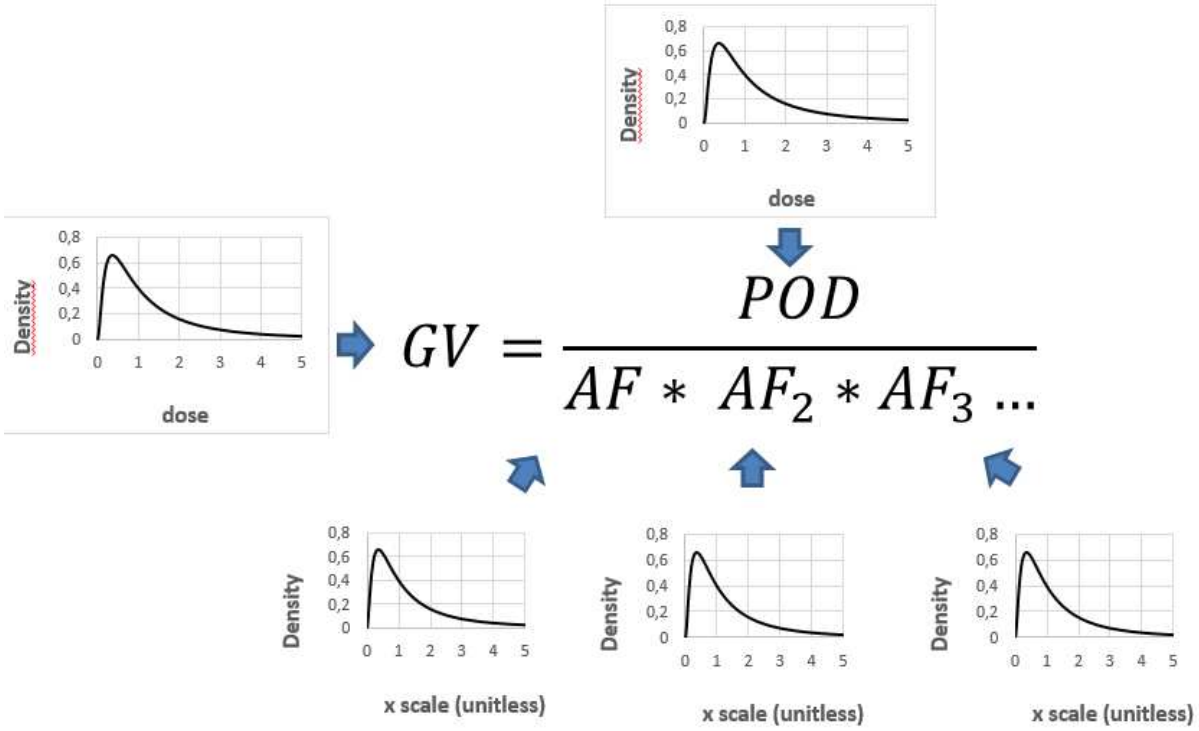


Figure 1-1 Combination of input distributions (displayed as probability density functions) of assessment factors (AF) to probabilistically derive a guidance value (GV)

The input parameters as well as the output guidance value are distributions. (Note that distributions can either be depicted as density functions, chosen here, with an integral of 1, or as cumulative distribution functions, with values between 0 and 1, see Annex 2).

The distribution obtained for the guidance value allows a better description and interpretation of uncertainties associated with the value – on costs of a mathematical more complex approach. Further, in order to apply a probabilistic approach several regulatory decisions are required:

- on the critical effect size or benchmark response (BMR), in order to determine the point of departure (which is ideally a benchmark dose)
- on the percentage of the target population to be covered by the value
- on the probability of achieving the above defined protection level.

This will be further analysed and discussed in this report.

In order to facilitate the understanding for readers new to probabilistic methods in Annex 2 the differences between empirical and parametric distributions are discussed as well as the presentation of distributions either as

- probability density function or
- cumulative distribution function.

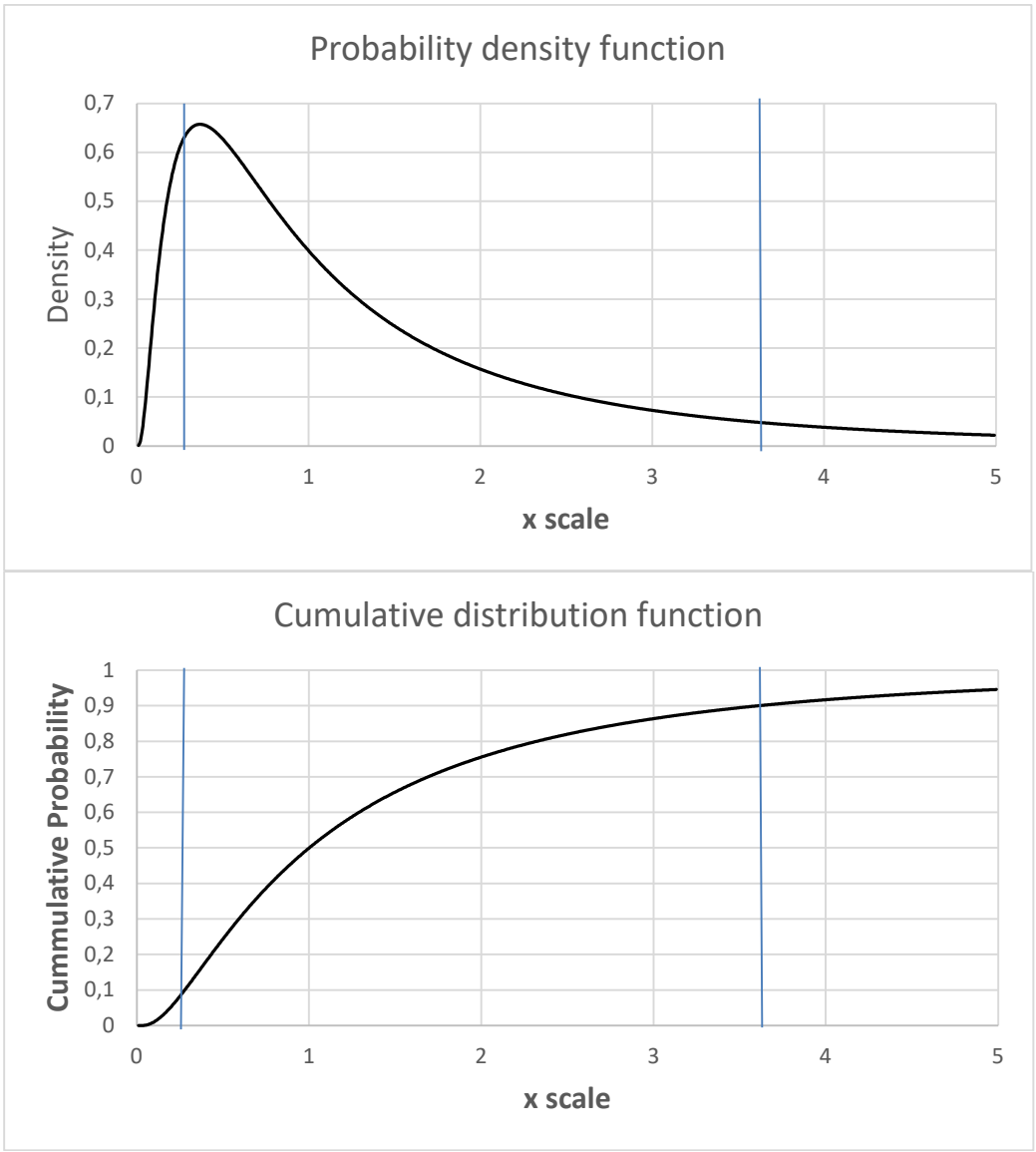


Figure 1-2 Probability density versus cumulative distribution function (with 10th and 90th percentiles as blue lines)

Further, some recurrent terms such as Monte Carlo simulation, bootstrapping, Bayesian methods, Markov Chain and Latin-Hypercube sampling are explained in Annex 2). But it is important to note that the tools presented here are able to be applied with a limited understanding of the mathematical background of the tools.

2 Historical development and existing proposals

2.1 Historical perspective

First approaches to probabilistic hazard assessment were published by authors in the USA (Baird et al. 1996; Baird et al. 2001; Price et al. 1997; Swartout et al. 1998). These approaches in the first instance tried to describe traditionally used assessment factors in the form of distributions. Hattis et al. described an approach for a probabilistic hazard assessment using empirical data for establishing distributions (Hattis et al. 2002). Weight was laid on the evaluation and use of the Hattis database on inter-individual differences in susceptibility. The results were compared with (then) existing RfD values from the IRIS database. This approach was further extended, again with using data from the Hattis database to describe both the toxicokinetic and toxicodynamic part of inter-individual variability (Hattis and Lynch 2007).

Scientists from the Dutch RIVM were the first to derive distributions from empirical data on the size of assessment factors (Slob and Pieters 1998; Vermeire et al. 1999). In several publications this approach was further developed and applied to example substances (Bosgra et al. 2005; van der Voet and Slob 2007; van der Voet et al. 2009). The distribution for intraspecies extrapolation in this work is based on theoretical considerations, not on an empirical database. The concept was further expanded to propose an integrated approach for probabilistic modelling of both hazard and exposure assessment (Slob et al. 2014; van der Voet et al. 2009). This integrated approach was also used for assessing carcinogens. For this work the MCRA software (https://mcra.rivm.nl/), which was developed for exposure assessment in the food safety area, was extended to cover hazard aspects. The MCRA software is not publicly available.

Hattis and colleagues, but also others emphasize that probabilistic methods would allow to harmonize approaches for threshold and non-threshold (carcinogenic) substances (Hattis et al. 2002; Slob et al. 2014).

In a research project for BAuA, FoBiG developed a probabilistic framework for deriving OELs. Based on empirical data, distributions for assessment factors were derived by statistical parameter fitting. These were combined with distributions for benchmark doses in Monte Carlo analyses (Schneider et al. 2004; Schneider et al. 2006). With several examples the authors also showed how to use substance-specific data to derive and use distributions for chemical-specific adjustment factors (CSAF) in the probabilistic assessment.

A few years ago the US National Research Council requested the US Environmental Protection Agency (US EPA) to develop probabilistic methods for deriving guidance values (NRC 2009). This fuelled the discussion on suitable approaches and pros and cons of probabilistic assessment procedures (Cooke 2010; Crump et al. 2010; Goble and Hattis 2010; Simon et al. 2016).

In 2015 Chiu of the US EPA and Wout Slob of RIVM proposed a probabilistic framework (Chiu and Slob 2015), which is also the core of a WHO report, developed in the frame of the IPCS: *International Programme on Chemical Safety* (IPCS) - Harmonization Project (WHO 2014). The approach presented there, which also includes an EXCEL®-based easy to use approximation to probabilistic modelling called APROBA, will be subject of a more detailed analysis in the following chapter. APROBA was further extended (APROBA-Plus) to cover also probabilistic exposure modelling (Bokkers et al. 2017).

The APROBA tool (for more information see chapter 3.1) was implemented as a web-based platform by Chiu and colleagues at two sites (Chiu et al. 2018)

- at https://wchiu.shinyapps.io/APROBAweb/

- and as part of the Bayesian Benchmark Dose (BBMD) Analysis tool platform by Shao and Shapiro (2018) (at https://benchmarkdose.com).

The BBMD platform was discussed with its dose-response modelling functionalities in our respective report on dose-response modelling. The probabilistic assessment tool added by Chiu and colleagues refers to the work by Chiu and Slob (Chiu et al. 2018; Chiu and Slob 2015) and is based on the approach proposed in the WHO/IPCS report.

At both platforms all distributions are assumed to be lognormal as it is done within the APROBA tool. There are differences between the two platforms. For example, no distinction is made between continuous and quantal endpoints in BBMD, whereas APROBA web seems to be an exact implementation of the functionalities of the EXCEL® tool, but with different forms for providing input and for presenting results. No further guidance is available for the BBMD implemented form. The APROBA web and the EXCEL® tool are further discussed in chapter 3.

Oldenkamp et al. provided two example probabilistic evaluations for the pharmaceuticals ciprofloxacin and methotrexate, based on the WHO/IPCS framework (Oldenkamp et al. 2016). Continuous endpoints were modelled for both substances. The main difference to the IPCS approach is that inter-individual variability is modelled in two steps, describing first the differences between the median individual in the general population and the median individual in the (more susceptible) subpopulation and in a second step the variability within the subpopulation. Secondary uncertainty due to limitations of the dataset on inter-individual variability is included in the model. The authors also included substance-specific data for informing the input distributions and compared results with the default distributions as proposed by WHO/IPCS (WHO 2014). As no substance-specific data on inter-individual differences in toxicodynamic could be found, data from the literature (Renwick and Lazarus 1998) were used to derive a distribution for this step.
The authors modelled the distribution for the average individual as well as the one for covering 99% of the whole population. The difference between the medians of these two distributions was interpreted as a measure for inter-individual variability, whereas the width of the distribution of the 99% distribution was taken as a measure for the uncertainty of the health-based value (the latter being much larger than the former).

Similar results were obtained when using substance-specific distributions or default distributions as proposed by WHO/IPCS (WHO 2014).

Chiu et al. tested the feasibility of the WHO/IPCS approach and its implications by performing (approximated) probabilistic assessments for a large set of substances and comparing it with RfD values of the US EPA (Chiu et al. 2018). In this comparison no benchmark doses were available, a fact that contributed much to the overall uncertainty. The values for 1% incidence at a 95% confidence level were found to be within one order of magnitude of the RfDs.

2.2 The WHO-IPCS document

2.2.1 Definition of "Target human dose" depending on data type

The WHO/IPCS Harmonization Project Document 11 is titled "Guidance document on evaluating and expressing uncertainty in hazard characterization" (WHO 2014). It focuses on addressing uncertainty and variability in health-based guidance values by probabilistic methods. The following terms are important to understand the approach in this document:

"target human dose", $\mathbf{HD_M^I}$

where

- HD is the reference or health-based guidance value resulting from the risk characterisation
- I is the fraction *I* of the population experiencing an effect (at dose HD)
- of magnitude (or severity) *M* or greater (for the critical effect considered).

In addition to I and M, the coverage (also called confidence or probability of effects) needs to be defined. Typically, a 95% probability is used to describe the outcome of the assessment, which means that there is a 5% probability that effects are more severe at HD than described by I and M.

I is the fraction of the general population not covered by HD. Typically, 1% is proposed in this tool, but a range from 0.1% to 5% is mentioned.

M is the critical effect size (or benchmark response, BMR) associated with the point of departure used for the assessment. M is set during dose-response modelling. For defining M, a distinction is made between different types of effect data:

- **continuous data**

M characterises the effect size at the boundary to adversity (and is equal to the BMR set when applying dose-response modelling)

Example for defining a HD for continuous data:

HD_{05}^{01} (for critical effect “reduction in red blood cell count”): the human dose at which 1% of the population shows a decrease in red blood cell counts of 5% or greater.

- **quantal deterministic data**

A distinction is made in the report between quantal data with an underlying continuous effect (e.g., histopathological effects in the liver, following (continuous) physiological changes) versus real stochastic effects (e.g., tumours or malformations). This will be discussed in more details further below.

Example for defining a HD for quantal deterministic data:

HD^{05} (for critical effect liver lesions): the human dose at which 5% of the population shows liver lesions.

- **quantal stochastic data**

Example for defining a HD for quantal stochastic data:

HD_{05}^{01} (for critical effect = risk of malformations): the human dose at which
1% of the population shows an individual extra risk of malformations of 5% or greater.

All these HD values are also a function of the probability (coverage) chosen: If a definite percentile of the probability function for HD_M^I is chosen, then HD_M^I would assume a single value (which, of course, would be different if a 90^{th} or a 95^{th} percentile is chosen).

The WHO/IPCS document describes three different approaches.

2.2.2 Non-probabilistic approach

In this approach, the lower and upper bounds for each hazard characterization aspect are combined by multiplication. Lower and upper bounds are typically chosen as 5^{th} and 95^{th} percentiles of uncertainty distributions, meaning, e.g., that the BMDL is divided by the 95^{th} percentiles of the distributions for interspecies, intraspecies and time extrapolation. Not surprisingly, this approach leads to wide ranges, spanning several orders of magnitude, between lower and upper bound estimates of the hazard value. This approach is of limited practical value.

This non-probabilistic approach is included – for comparison – in the APROBA EXCEL® spreadsheet.

2.2.3 Approximate probabilistic approach (APROBA)

In the approximate probabilistic approach all uncertainties are assumed to be lognormally distributed and are described by independent lognormal probability distributions. With this assumption, algorithms can be implemented in an EXCEL® spreadsheet and can be solved numerically without Monte Carlo simulations.

The APROBA EXCEL® spreadsheet contains default distributions (see chapter 2.2.5), which can be changed by the user (see chapter 3.1 for more information).

2.2.4 Full probabilistic approach

With this approach, uncertainty distributions of any form are combined probabilistically. As the type of distributions is not restricted to lognormal distributions, a mathematical exact solution is not possible. Calculations need to be done using Monte Carlo simulations.

A full probabilistic assessment is more flexible, can use any kind of distribution and can include additional aspects of uncertainty. But it can be run with the same distributions as used in APROBA and – according to the report - would give similar results.

For a full probabilistic approach there are no ready-to-use tools described in the report.

2.2.5 Uncertainty distributions

2.2.5.1 Point of departure

Although the BMDL is given clear preference, also NOAEL and LOAEL are foreseen as possible points of departure (POD).

When dose-response modelling is performed, upper and lower bounds of the BMD are obtained (BMDU and BMDL, resp.). A distribution of the BMD expressing the uncertainty of the BMD can be obtained either

- using the BMDL and BMDU and assuming a lognormal distribution (this approach is used in APROBA)
- approximating a distribution by the bootstrap method or by a Bayesian approach.

Note that the POD and consequently also the resulting $\mathbf{HD_M^I}$ is different for the different data types with regard to the effect size chosen:

Continuous data:
$BMDL_M$ with M being the BMR chosen for a specific continuous endpoint is the POD. As the data points used for dose-response modelling are the group averages, it can be assumed that each data point and also the BMDL stands for an average response rate of 50% within each group with regard to the chosen M.

Quantal deterministic data:
Although this does not become clear from the WHO/IPCS report and the description of the APROBA tool, it is proposed here to use the $BMDL_{50}$ as the point of departure (personal communication Wout Slob, 21 May 2019). Implications will further be discussed in chapter 4.

Quantal stochastic data:
For stochastic endpoints the $BMDL_{10}$ is proposed as POD, in agreement with current practice.

Also for using NOAEL and LOAEL values, approximate uncertainty distributions are derived and implemented in APROBA. These uncertainty distributions are wider than typical BMD distributions.

2.2.5.2 Exposure duration

The following distributions are proposed for exposure duration extrapolation:

Subchronic to chronic extrapolation, based on the evaluation of Bokkers and Slob (2005):

Distribution with GM = 2, with P95/P50 = 4 [(P05, P95) = (0.5, 8)]

Subacute to chronic extrapolation, based on various evaluations of NOAEL ratios in the literature up to the year 2011:

Distribution with GM = 5, with P95/P50 = 8 [(P05, P95) = (0.625, 40)].

The database on exposure duration extrapolation will be discussed in detail in a separate report of this project.

2.2.5.3 Interspecies extrapolation

Interspecies extrapolation is performed in two steps:

- adjustment of the dose for differences in body size between test animal and humans by allometric scaling
- accounting for potential and unknown (chemical-specific) differences between species in toxicokinetics and/or toxicodynamics.

Based on evaluations in the literature on quantitative species differences the following distributions were proposed:

For step 1:

An allometric scaling exponent of 0.7 is used; the distribution is established based on the 95% confidence interval of the exponent of 0.66 – 0.74.

For step 2:
TK/TD uncertainty after accounting for body size differences, based on evaluations in the literature on quantitative species differences the following distributions:

Distribution with GM = 1, P95/P50 = 3 [(P05, P95) = (1/3, 3)]

Again, the underlying data on interspecies extrapolation will be discussed in a separate report.

For inhalation exposure the report assumes as a default that in the central tendency there are no differences between animals and humans with regard to deposition of particles or doses of gases (i.e. median of ratios between animals and humans is 1). A low uncertainty with a 95th percentile to median ratio of the lognormal distribution of 2 is assumed, but no data are presented to support these assumptions. But this assumption is not used in the APROBA tool, where the same distribution (see above) for oral and inhalation studies is used.

2.2.5.4 Inter-individual variability

The distributions for inter-individual variability are derived from literature evaluations, mainly from Hattis and co-workers (Hattis et al. 2002; Hattis and Lynch 2007) and Renwick and Dorne (Dorne et al. 2005; Renwick and Lazarus 1998).

The derived distributions cover the whole population (no distinction is made, e.g., between adults and children, as the differences were not statistically significantly different). Separate analysis and distributions are presented for toxicokinetics and –dynamics, but – assuming independency – the distributions can also be combined to one. A generic lognormal distribution results, which is characterised by log(GSD_H), which represents the distribution of the spreads of groups of individuals with differing susceptibility found in the literature. The log(GSD_H) is characterised by the following values: median: 0.324; 95th percentile 0.697. How geometric standard deviations are used to describe variability in human populations will be discussed in more details in another report (on intraspecies extrapolation).

To exemplify the meaning of this distribution, it can be translated into assessment factors (ratio between the equipotent doses in average and susceptible humans):

$AF_{Intra\text{-}I}$ = Factor covering (1 – I) of the population = $GSD_H^{Z1\text{-}I}$,

For *I* = 5%, 1% and 0.1%, the corresponding values for $z_{1\text{-}I}$ are 1.6449, 2.3263 and 3.0902.

Table 2-1 Factors for intraspecies extrapolation

chosen incidence I	**5th perc**	**50th perc.**	**95th perc**
5%	1.8	3.41	14
1%	2.2	5.67	42
0.1%	2.9	14.23	143

Interpretation for a chosen incidence level of 1%: a factor of 5.67 is required to cover 99% of the population with a probability of 50%. At the 95% probability level a factor of 42 is required to cover 99% of the population.

With this concept of having individual distributions for covering certain percentiles of the target human population the WHO/IPCS report follows the methodological approach as developed in the BAuA project (Schneider et al. 2004; Schneider et al. 2006).

2.2.5.5 Route-to-route extrapolation

The uncertainties of route-to-route extrapolation are discussed, but no default distribution is proposed.

2.2.5.6 Other uncertainties

The APROBA tool as well as the full probabilistic model are open for adding additional uncertainty distributions (in case of APROBA up to 3, all in the form of lognormal distributions). Examples are uncertainties due to differences in exposure patterns (e.g., 4 hours daily inhalation exposure instead of 8 or 24 hours) or the uncertainty introduced by an incomplete database.

2.2.5.7 Secondary uncertainties

Secondary uncertainties are uncertainties associated with the used distributions themselves. For example, there is uncertainty in the data used for deriving the ratios between studies of different exposure duration, e.g., the NOAELs used for calculating NOAEL ratios. With regard to inter-individual differences in susceptibility, the individuals studied and/or the chemicals under study might not be representative for the target population and/or the chemical under assessment.

Secondary uncertainties are considered in some probabilistic models (Oldenkamp et al. 2016). They cannot be assessed in APROBA. They are generally considered to be substantially lower than primary uncertainties.

3 Tools for risk assessment

3.1 APROBA and related web tools

The APROBA ("Approximate Probabilistic Analysis) tool was developed in the frame of the WHO/IPCS "Guidance document on evaluating and expressing uncertainty in hazard characterization" (WHO 2014). In its original form it is EXCEL®-based, but two web-based forms have been made available by Chiu and colleagues (Chiu et al. 2018) at:

- https://wchiu.shinyapps.io/APROBAweb/
- and as part of the Bayesian Benchmark Dose (BBMD) Analysis tool platform by Shao and Shapiro (2018) (at https://benchmarkdose.com).

The EXCEL®-form and the APROBAweb-version are discussed in more details. The BBMD version does not have the full functionality and no additional features.

The EXCEL® version has the following main features:

- APROBA is specifically designed for probabilistic hazard assessment
- as an EXCEL-based tool it works on the simplification that all distributions are lognormally distributed; this allows an analytical solution of the algorithms
- three types of dose-response data are discerned and dealt with in slightly different form (see chapter 2.2):
    - continuous data
    - quantal deterministic data
    - quantal stochastic data.
- BMDs, LOAELs or NOAEL can be used as input data for the point of departure
    - the BMD is assumed to be lognormally distributed and the BMDU and BMDL are required to define the distribution
    - default uncertainty distributions for NOAELs and LOAELs are implemented.
- There are default distributions implemented for
    - uncertainties associated with allometric scaling
    - interspecies extrapolation
    - exposure duration extrapolation
    - inter-individual variability
- The tool allows to compare with a deterministic assessment
- A sensitivity analysis in tabular form allows to assess the contribution of individual factors to the overall uncertainty
- Results are presented graphically and in tables.

In principle, the web-based version contains all these features as well. Differences between the two versions are listed in the following table.

Table 3-1 Comparison of main features of the EXCEL® and the web-based form of APROBA

	APROBA EXCEL®	**APROBAweb**	**Remarks**
input data for BMD distribution	BMDL and BMDU	BMDL and BMD	
default values for body weight	e.g., rat: 0.4	e.g., rat: 0.3045	different body weight defaults are used in the two versions, but can be changed
distributions for exposure duration, interspecies extrapolation	input data required as lower and upper bound (5^{th} and 95^{th} percentile) values	input data required as central value (50^{th} percentile) and ratio 95/50 percentile	Example: sub-chronic to chronic defaults: xls: 0.5 / 8 web: 2 /4
distribution for inter-individual variability	input data required as lower and upper bound (5^{th} and 95^{th} percentile) values	input data required as GSD_H (50^{th} perc.) and log GSD_H	for interpretation of GSD_H see 2.2.5.4
graphical presentation	yes, plot of dose versus incidence, at various probability levels	yes, plot of dose versus incidence, at various probability levels (with normal and lognormal y scale); in addition PDF* of $HD_M{}^I$	extended graphical presentation in web tool
sensitivity analysis of uncertainties	in tabular form	in graphical form	
export and report generation	no, but excel file can be saved	only export as csv	

*Probability density function

The following figure shows the graphical presentation of results in the APROBA EXCEL® tool. On the x axis the dose is given (on a log scale). The y axis indicates the incidence or the percentage of the population covered (incidence 1% means 99% of the population is covered). The coloured lines give the relationships between dose and

incidence for selected probabilities (from left to right): 99%, 95%, 90%, 10%, 5%, 1%). For example, 99% probability here means that with a 99% percent probability the dose leading to the incidence on the y axis is not higher than the dose given on the x axis.

If incidence and probability are determined, then a fixed dose is obtained: HD_M^I (blue square). The black vertical line indicates where a deterministic value would be located, using default assessment factors. The red vertical line is shown, if an exposure estimate is entered.

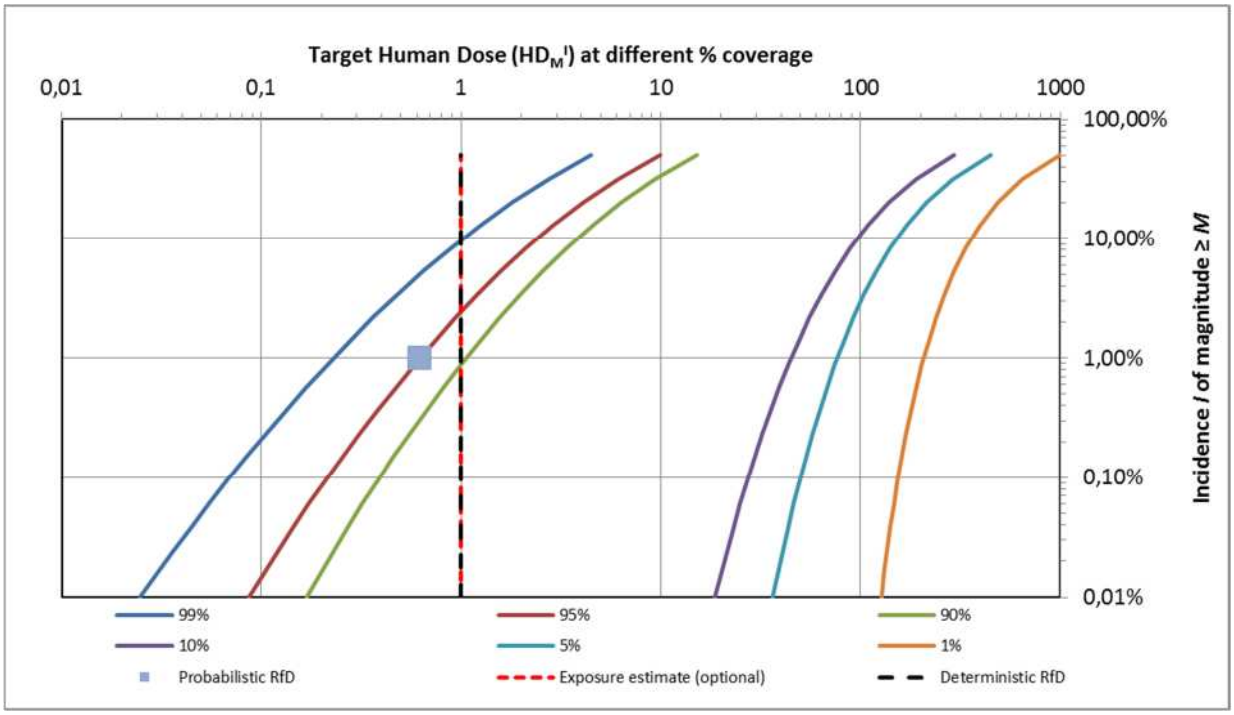


Figure 3-1 Presentation of results in the APROBA EXCEL® tool

The presentation of results in the web-based APROBA tool is very similar (see **Figure 3-2**). The incidence scale can be depicted either normal or on a logarithmic scale.

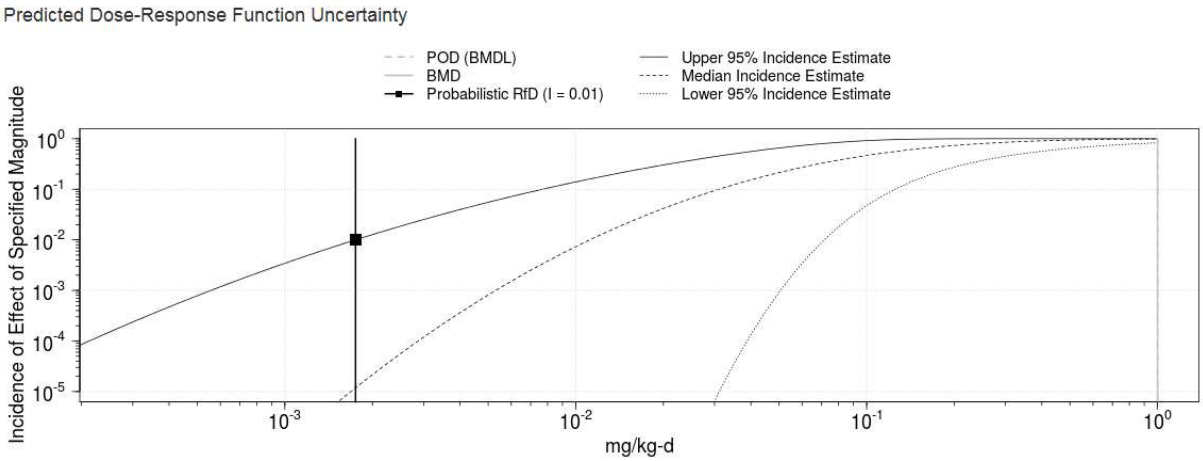


Figure 3-2 Presentation of results in the APROBA web tool

In addition, the web tool provides a probability density function of the resulting HDMI (with fixed incidence), describing the uncertainty of HDMI (note that the PDF takes the shape of a normal distribution, as the x axis (dose) is given in a log scale).

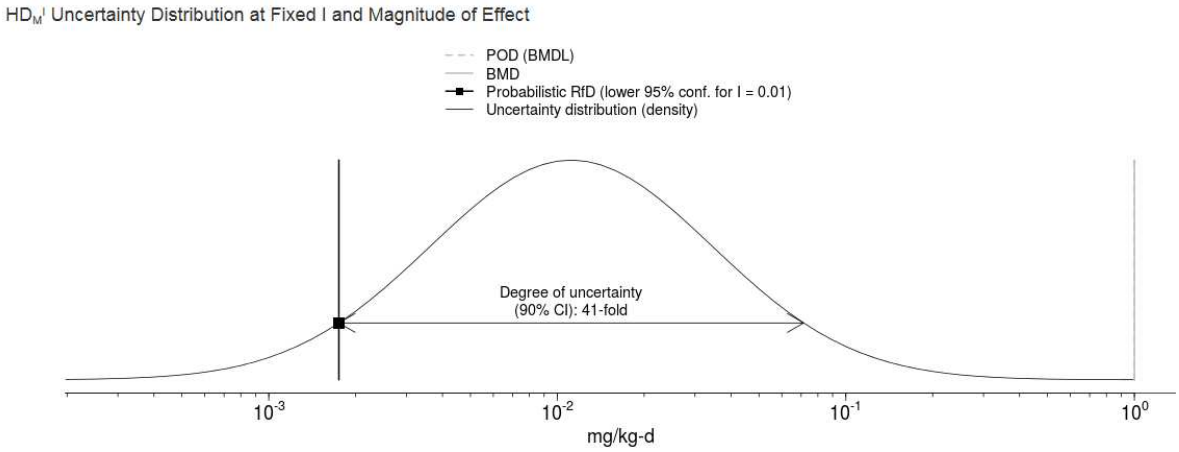


Figure 3-3 Presentation of the probability density function of HD_M^I in the APROBA web tool

3.2 EFSA's web-based tool

On behalf of EFSA a publicly accessible tool for Monte-Carlo simulations was developed respectively is under development (Seynaeve and Verbeke 2017). The tool is accessible (after simple registration) at EFSA's Statistical Models website[1]. Main features of the tool are:

- it is designed to be used for risk assessments (exposure, hazard, any kind), but it is not populated with default distributions

[1] https://r4eu.efsa.europa.eu/

- it allows to use predefined parametric distributions (with their parameter values), but it is also possible to enter distributions via a range of percentile values (various import functions are available for use of existing data; however, guidance needs to be developed for full usage)
- algorithms (model equations) can be entered by an easy-to-use formula editor
- two modes are available for Monte Carlo simulations: simple random sampling and Latin hypercube sampling
- there is a sophisticated presentation of results, in tabular and graphical form
- sensitivity analysis is performed as part of the routine and presented graphically
- input data and assessments can be stored, and output reports can be exported in pdf or word format.

The tool was initially developed for specific problems in food safety but is designed to be applied for any kind of question to be solved with a probabilistic approach. During its development it was compared and validated against @risk (see chapter 3.3), both with regard to model fitting and Monte Carlo analysis. A technical report describing the main features of the tool is available (Seynaeve and Verbeke 2017) and a short user manual is available at the website.

The following figures illustrate some of the main features of the tool.

Creation of distributions of input variables:
Parametric distributions of input variables can be determined by entering determinants like GM and GSD (for a lognormal distribution), but also, as shown in the figure below, by determining the type of distribution (e.g., lognormal) and entering percentile values determining the shape of the distribution.

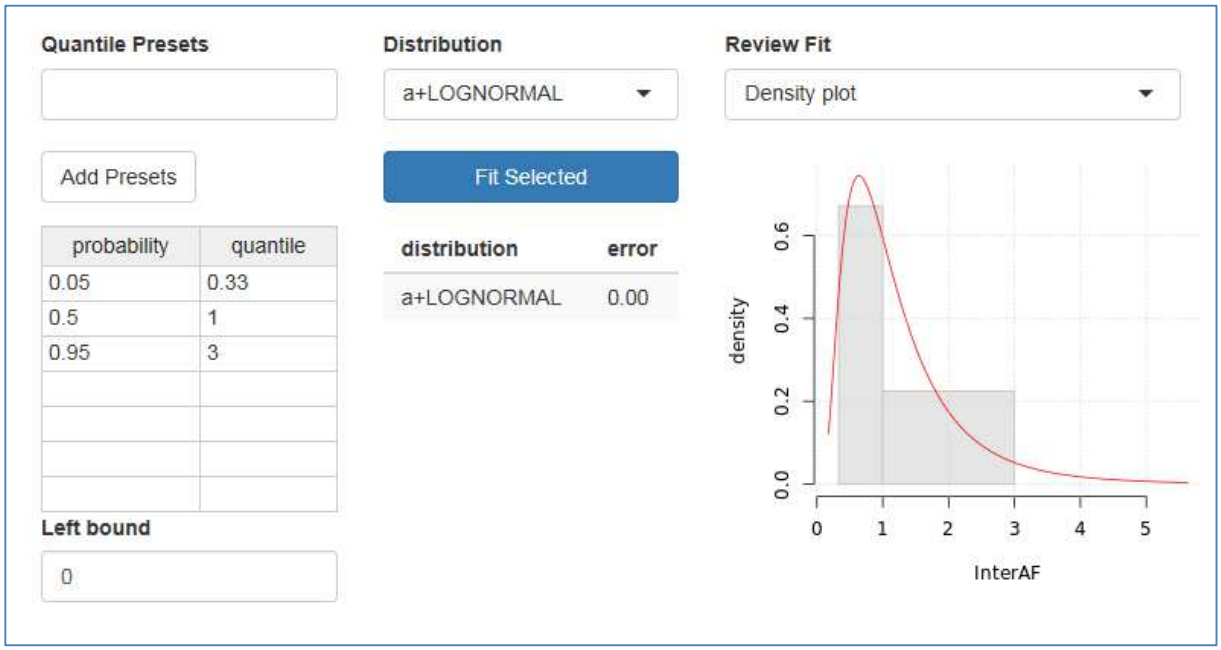


Figure 3-4 Creation of a lognormal input distribution by providing percentiles

Functionalities are planned to allow fits to empirical data, but these are not yet fully implemented.

Editor for model equations

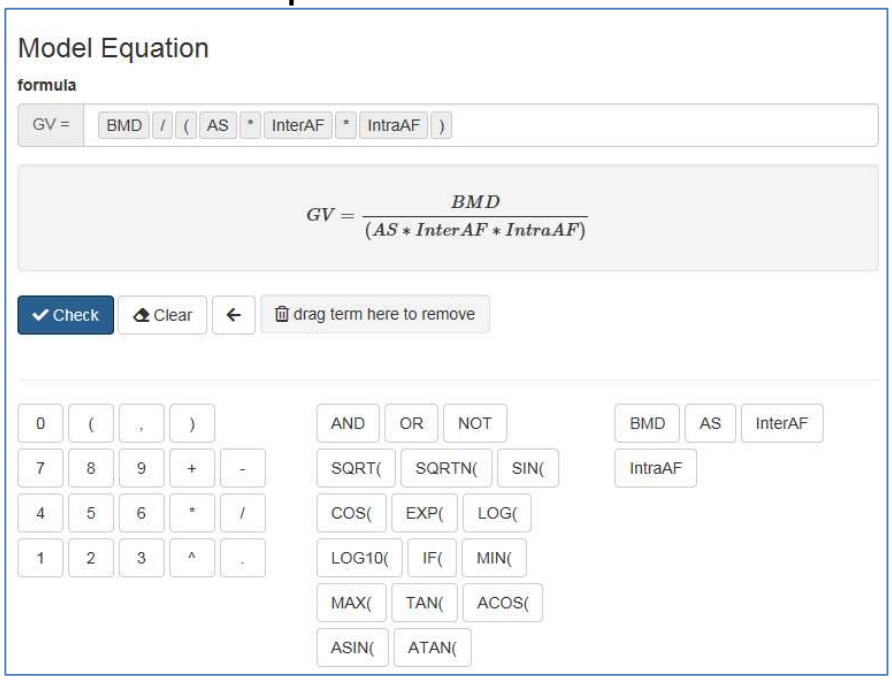


Figure 3-5 Editor for model equations

An easy-to-use formula editor is available in order to connect input and output variables by a defined algorithm.

Monte-Carlo simulation

Simulation conditions can be set in the respective module and the simulation can be followed in real-time.

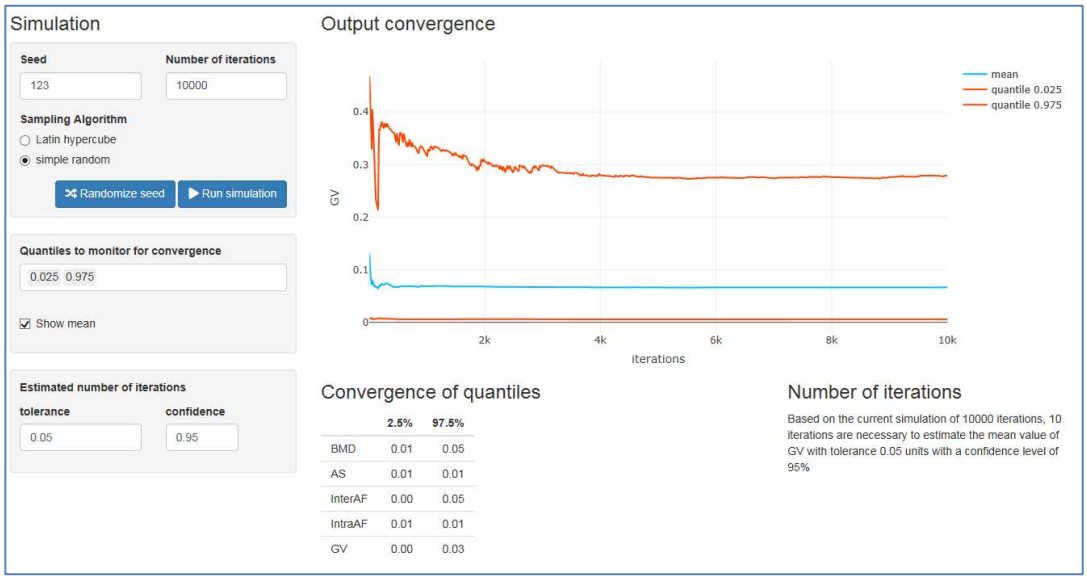


Figure 3-6 Simulation tool

Results presentation

The resulting distribution is given in tabular form (last row in table), together with all input distributions, as well as in graphical form as probability density function (PDF, left) and cumulative distribution function (CDF, right).

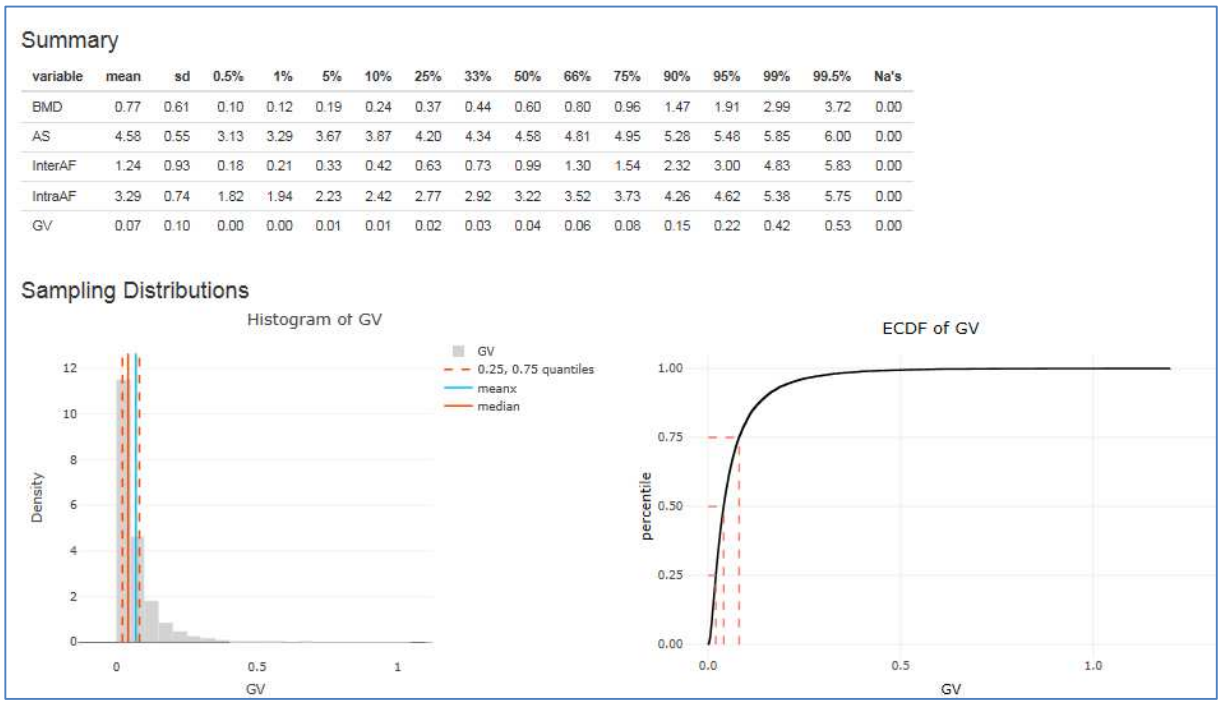


Figure 3-7 Presentation of results in the EFSA tool

Input distributions are displayed in a similar way. Results can be exported as pdf, docx, or csv file.

3.3 Other tools for full probabilistic assessments

Various software implementations exist to perform Monte Carlo Simulations. Commercially available add-in based tools are e.g. @risk, Crystal Ball, Risk Solver or Model Risk. These tools can be used for fitting models to empirical distributions and provide a large variety of uncertainty distributions as potential data input for Monte Carlo simulations. A Bayesian approach of sample generation (Markov chain Monte Carlo) is implemented in non-commercially and freely available software as WinBUGS, JAGS or STAN (Chiu and Slob 2015) or Nimble.

It should be noted that these software tools were not developed for the specific use in toxicology. Main application areas, for example of @risk are economics or insurance business. With the availability of the EFSA tool described there exist an easy-to-use alternative allowing a full probabilistic assessment.

4 Analysis and discussion of existing approaches

4.1 Principal approaches

The common principle in all probabilistic approaches published so far is to replace the input parameters of the algorithm

HD = POD/ (AF1 * AF2 * AF3 *..)

by distributions characterising uncertainty and/or variability in these parameters.

The POD typically used in probabilistic approaches is the distribution to the benchmark dose (BMD), although APROBA also allows using NOAELs or LOAELs (with predefined uncertainty distributions).

In early publications distributions were merely reflections of the default assessment factors used (Baird et al. 1996; Baird et al. 2001; Price et al. 1997; Swartout et al. 1998). Authors from RIVM then proposed data-derived distributions for exposure duration and interspecies extrapolation (Slob and Pieters 1998; Vermeire et al. 1999). Hattis and colleagues (Hattis et al. 2002) and later Schneider et al. (Schneider et al. 2004; Schneider et al. 2006) proposed empirical distributions also for intraspecies extrapolation. In principle, all systems allow to add additional uncertainty distributions and to replace empirical default distributions by substance-specific distributions for certain extrapolation steps.

All approaches published since 2001 use allometric scaling for accounting for differences in body size in addition to empirically derived distributions for various uncertainties. It can be concluded that there is a high agreement in the principal construction of probabilistic models for hazard assessment.

Slight methodological differences in how Monte Carlo simulations are performed (e.g., whether random or Latin Hypercube sampling is used) are not expected to lead to relevant differences.

With APROBA a tool is now available, which is based on the assumption that all distributions are lognormally distributed, which allows easy analytical solutions, which can be implemented in EXCEL®. The new EFSA tool, although still under development and without adequate guidance yet, is a ready-to-use tool for full probabilistic assessments. Practicality and adequacy of both tools is investigated with the two examples analysed in chapter 6.

4.2 Preconditions, pros and cons

Probabilistic methods are not yet often used for OEL setting. The model by Schneider et al. was specifically developed for the workplace, but is not used on a regular basis

(Schneider et al. 2006). ANSES is experimenting with a probabilistic approach, in which distributions for assessment factors are numerically set to represent default values used previously (Vernez et al. 2018). ANSES combines the probabilistic approach with expert decisions and no rule is mentioned by Vernez et al. for the coverage of the exposed population. For the two examples presented, OELs were around the 9th percentile of the resulting probability density distribution.

In order to perform a probabilistic assessment to derive a definite OEL value, principal regulatory decisions need to be taken:

- on the level of adversity by determining a benchmark response (BMR)
- on the percentage of the whole population, which should be covered by the OEL
- on the coverage or confidence of the assessment by determining the probability level (e.g., 95% percentile of the probability distribution).

If the first two parameters are set, then a distribution is obtained, which describes the probability at a given dose of not exceeding the chosen effect size and incidence. If all three parameters are set, a single dose value is obtained at which with at a selected confidence level (e.g., 5th percentile of the distribution for 95% confidence) the determined critical effect size (BMR) is not exceeded at the chosen incidence (e.g., in 99% of the target population).

Although assessors would prefer to derive “definitely” safe levels, it is obvious that uncertainties and probabilities are associated also with deterministically derived OELs. The main difference is that with deterministic OELs these uncertainties are hidden. Probabilistic methods provide more information on the characteristics of the outcome of the assessment. By explicitly stating the percentile of the population covered, the misinterpretation that the OEL is safe for everybody in a population of individuals with diverse susceptibilities is avoided (Hattis et al. 2002). On the other hand it can be argued that the complex presentation might imply a (too) high precision (Goble and Hattis 2010).

Advantages of probabilistic methods:

- It requires responsible persons or bodies to make conscious choices on the aim of the assessment and the protection objectives of the OEL
- These choices are made transparent
- Uncertainty and variability are quantified
- Risk management receives information on the probability of adverse effects in the chosen part of the population in a three-dimensional matrix of dose, percent of population and probability of effects
- This matrix allows for conclusions on effects at dose levels above the OEL (see next chapter)
- These methods allow for a sensitivity analysis, which helps to focus resources on the most relevant sources of uncertainty.

Disadvantages:

- Probabilistic modelling requires a principal understanding of the approach and the interpretation of the results by risk managers, stakeholders etc.

- Its complexity might cause acceptance problems, requires education and may lead to a reduction in transparency (if not understood)
- It is more resource-intensive
- Ready-to-use tools (APROBA, EFSA tool) have been developed only recently; they need proper testing and introduction and use experiences need to be gathered in a larger community.

4.3 Uncertainty versus variability, secondary uncertainty

As already explained above, the inherent uncertainty of the BMD and values for interspecies extrapolation, exposure duration extrapolation and others is described and quantified by their distributions:

- The BMD distribution describes the uncertainty and variability observed in an experimental study under the given conditions, i.e., the uncertainty about the observed BMD being the “real” BMD.

- The distribution for interspecies extrapolation quantifies the uncertainty of not knowing whether for the assessed substance the differences in toxicokinetics and –dynamics between humans and experimental animals is larger or smaller than the central tendency of the empirical database of chemicals.

- The same holds true for exposure duration extrapolation: the distributions describe the ratios (chronic to shorter-term studies) found for a large set of substances; the uncertainty stems from not knowing which factor is the correct one for the substance in question.

Further uncertainty distributions can be added where required (e.g., to describe uncertainties in route-to-route extrapolation, uncertainties due to an incomplete database, or uncertainty introduced by applying read-across concepts). Also, adjustments of the POD, for example for differences in the exposure scenarios of the animal experiment compared to the workplace situation, may contain uncertainties, which might be approximated by distributions.

The intraspecies distribution describes the variability in susceptibility between individuals in the target group, but also contains aspects of uncertainty: the distributions are derived from individual datasets on susceptibility differences between individuals of small groups (e.g., volunteers of a toxicokinetic study of a pharmaceutical) from a certain range of chemicals (many of them pharmaceuticals). Such datasets might underestimate the variability in the target population because the groups of volunteers (younger, healthy adults) and/or pharmaceuticals might not be representative for the more variable entity of the workforce and chemical substances used at the workplace (Hattis et al. 2002).

Differentiation between uncertainty and variability (within the worker population) might be simply achieved approximately by omitting/adding the intraspecies distribution, although (as discussed above) this distribution contains some aspects of uncertainty as well.

Some approaches (e.g. Oldenkamp et al. 2016) also include aspects of secondary uncertainty. Secondary uncertainties result from lack of knowledge on whether the chosen distributions are actually the right ones. For example, inter-study variation is included in the BMD distribution to a limited extent and repeating the experimental study several times might lead to a different BMD distribution. As mentioned above, the datasets used for describing inter-individual variability might not be representative with regard to the human individuals participating in those studies and the chemical substances used. This leads to secondary uncertainty not included in the model when using the distribution derived in the first place.

4.4 Types of dose-response data

Dose-response modelling typically distinguishes between continuous and quantal data (see separate report on "Dose-response modelling", Part 2). The WHO/IPCS report further divides data presented in quantal form into quantal deterministic and quantal stochastic data (WHO 2014), with consequences for the handling in the APROBA tool and for the interpretation of the resulting "target human dose", HD_M^I**.**

The interpretation of HD_M^I in the case of continuous data is that a part of the target population, which is defined by the incidence I (e.g., 1% of the target population) can experience an effect defined by M or higher.
For example, if 5% increase in thyroid weight is chosen as M, then at or above dose HD_M^I (with I=1% and the chosen probability) 1% of the population will suffer from a ≥5% increase in thyroid weight. The POD used here, the $BMD_{05\text{-thyroid weight}}$ represents the dose, at which the <u>average</u> group response in the experiment was 5% increase in thyroid weight.

For quantal data according to the WHO/IPCS document the assessor needs to decide whether the quantal effect is a "real" stochastic effect or whether there is an underlying (unknown or unreported) continuous effect. For the former type of effects, a POD of 10% extra risk is proposed. In contrast, for the latter the $BMDL_{50}$ is proposed by the authors as a POD (personal communication with Wout Slob, 19 May 2019 and Weihsueh Chiu, 30 May 2019), because

- the difference between the BMDL50 and the BMDL10 reflects only the variability in the experimental animals, which is not of interest
- the underlying continuous effect, if measured, would be provided as group averages; hence, the quantification would not be based on the most sensitive animal
- the BMDL50 would better correlate with the effect level related to the critical continuous effect.

The relationship between the data types is shown in the following figure. This theoretical dataset consists of the individual responses of 5 animals in each of 4 dose groups (including control). The effect measured is weight of organ X, measured in g. In order to transform these continuous data into quantal data, a cut-off needs to be defined. Here, 10 g organ weight is chosen: any individual animal showing a weight of organ X below 10 g is considered affected.

Table 4-1 Theoretical dataset – organ weights of animals in 4 dose groups

Dose (mg/kg bw/day)	Organ weight of individual animals (g)					Arithmetic mean (g)	Incidence (%)
0	17	11	14	15	13	14.0	0
10	14	8	11	16	12	12.2	20
30	9	11	12	14	6	10.4	40
100	5	7	11	8	6	7.4	80

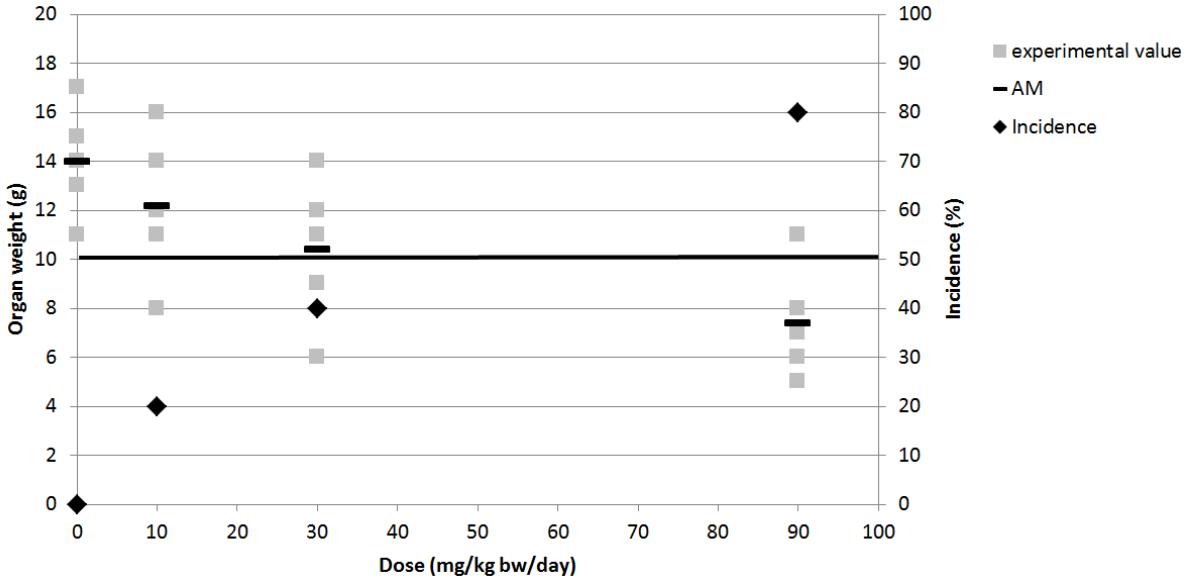


Figure 4-1 Dose-response data for continuous effect data and the transformed quantal data (bold line: adversity criterion: 10 g organ weight)

This example shows that indeed transformation of a set of continuous data into quantal data leads to a lower BMD(L), as the quantal BMD is based on the most sensitive animal(s): the BMD_{10} can be expected at approx. 5 mg/kg bw/day, whereas the BMD for the critical effect 10 g organ weight is close to 30 mg/kg bw/day.

As inter-individual variability in the human population is addressed separately, it can be argued that the variability between animals does not need to be considered. The $BMD(L)_{50}$ is expected to be much closer to the $BMD(L)_{critical\ continuous\ effect}$ than the $BMD(L)_{10}$. This problem was also already addressed in the BAuA research project (Schneider et al. 2004).

In contrast to these quantal deterministic data with underlying continuous effects, the authors assume that there can be real stochastic effects (i.e. effects, which, upon repetition realise not with certainty, but only with a certain probability; examples: tumour formation, malformations), for which not different susceptibilities of individual

animals causes the variability, but the random stochastic event. In that case they propose to use the $BMDL_{10}$ as POD.

The practical consequences of this differentiation need further discussion. Whereas it is likely that many or most quantal effects fall under the quantal deterministic category (i.e., it is likely that there are some measurable physiological or pathogenic changes occurring as the basis of the quantal effect), it is difficult to decide upon this in individual cases. Also, deviating from the standard procedure of using the $BMD(L)_{10}$ for quantal effects in dose-response modelling certainly would need a broader discussion and consensus in the risk assessor community. Further, examples are required to demonstrate the practical and numerical consequences.

In the current situation assessments with quantal data, if carried out probabilistically, should use the $BMDL_{10}$ as POD to allow comparing it with the traditional deterministic approach. But wherever feasible, in addition, $BMDL_{50}$ should be used as POD to gather experience and to discuss and compare the approaches.

4.5 Practicality

With APROBA (EXCEL® or web-form) and the EFSA web-based tool there are now easy to use (approximate in the case of APROBA) probabilistic tools available, which allow for a fast and easy application. The EFSA tool is under development and proper guidance documentation is still required. A short manual is available at EFSA's website (https://r4eu.efsa.europa.eu/). Nevertheless, for probabilistic assessments to be properly performed, a good understanding of principles, objectives and rules by the assessor is required.

APROBA is restricted to using predefined lognormal distributions for describing assessment factors. In contrast, the EFSA tool allows fitting models to empirical datasets or to use other parametric distributions and therefore is broader in its applicability. However, no predefined default distributions are given. Therefore, more expertise is required for the EFSA tool in defining the distributions. A detailed guidance document (expected to come soon) is still lacking.

Both tools come with an easy-to-use sensitivity analysis, which indicates which input distribution contributes most to overall uncertainty. This is very helpful asset for every assessment. (But note that differences were observed in the sensitivity results obtained with the examples in chapter 6, which need further analysis.)

With these available tools the next stage in probabilistic modelling can be achieved. This would consist in probabilistic assessments performed in parallel to deterministic ones for comparison and discussing pros and cons and possibilities for improvements of the approach. Nevertheless, it is expected that the use of probabilistic tools will increase only slowly. A main reason definitely is the higher complexity of probabilistic approaches. Risk assessors need to be familiar with the tools, its advantages and disadvantages. Probabilistic approaches might be especially useful under specific circumstances. Several applicability cases can be discerned for a substance, for which an OEL needs to be derived:

- Case 1: Uncertainty is assumed to be high, but exposures are orders of magnitude below the deterministic OEL

- Case 2: Uncertainty is low (e.g., OEL is based on qualitatively good human data), and exposure is in the range or above the OEL

- Case 3: Uncertainty is high, and exposure is in the range or above the OEL.

In the first case there is not much need for refining the OEL by probabilistic methods, as it would not have practical consequences in a situation considered to be safe. In the second case sufficient knowledge is available to conclude on adequate risk management measures. Case 3 is the one, which would benefit from a probabilistic approach:

- the sensitivity analysis would indicate major sources of uncertainty (which might be reduced by additional efforts before investing into costly risk management measures)
- the derived probabilistic OEL would indicate the confidence the assessor can have when using the value
- the assessment would inform about the likeliness of adverse effects at the determined exposure levels (see following chapter).

5 Description of risks above the OEL

The derivation of OELs aims at identifying a concentration, below which adverse effects are unlikely to occur (see separate report on the “Comparison of methods” for details and differences in the definition of existing approaches). But for certain situations and regulatory problems it is also important to have information on the likelihood and incidence/severity of effects above the OEL. Examples for such problems are

- the OEL is exceeded in a certain real situation and adequate risk management measures need to be defined
- an impact assessment is performed to identify the consequences of setting an OEL at a certain level.

An example for the latter is the setting of binding OELs by the European Commission for carcinogens under the Carcinogens and Mutagens Directive (Directive 2004/37/EC). Although this is typically performed for non-threshold carcinogens, recently an impact assessment was carried out for substances with a presumed threshold (nickel compounds, benzene, acrylonitrile).

The dose-response modelling allows to predict the incidence (for quantal data) or the severity (in terms of the effect size in the case of continuous data) of effects above the benchmark dose, these predictions apply strictly only to the test system and test conditions applied to gain the dose-response data, typically the experimental setting of the toxicity study used (see separate report on “Dose-response modelling”). Only where adequate long-term human data are used, a direct conclusion may be derived for the situation in humans at higher doses.

The result of a probabilistic assessment is a three-dimensional matrix with the dimensions dose (or concentration), incidence and probability. The consequences of increasing the dose for the expected incidence of effects at a given probability can directly be derived from the obtained distribution. Especially, the graphical presentation in APROBA allows to rapidly conclude on the increase in incidence with increasing dose (see examples in chapter 6):
For Example 1 (renal tubule hyperplasia induced by 3-MPMD (Annex 1) the following result was obtained:

HD_M^I = 0.0011 mg/kg bw/day, with
M: extra risk of 10 percent for renal hyperplasia
I: 1% of the population
which indicates that at this dose – with a 95% confidence – 1% of the population will have an extra risk of 10% for renal tubule hyperplasia (which can be roughly approximated to a 0.1% extra risk for that effect in the population)

At a 10fold dose at the same confidence level of 95% the incidence level would be approx. 25% for a 10% extra risk.

6 Conclusions

Two recent developments in the area of probabilistic hazard assessment are described in this report:

The APROBA tool developed in the frame of the WHO/IPCS project on “Evaluating and Expressing Uncertainty in Hazard Characterization” (WHO 2014) is an EXCEL®-based tool allowing approximations to full Monte Carlo analyses by using lognormal input distributions for all parameters.

The Monte Carlo tool developed by EFSA, currently under development at EFSA, allows for full Monte Carlo analyses, including distribution fitting and use of various kinds of distributions.

In view of these new developments use of probabilistic approaches to hazard assessment are simplified and their use for

- method development and discussion of (combination of) deterministic factors
- comparison with standard assessments using deterministic factors
- refined assessments of complex cases

is encouraged.

References

Aral, M.M. (2010). Uncertainty and variability analysis. In: M.M. Aral (Ed), Environmental Modeling and Health Risk Analysis (Acts/Risk), pp. 333-355.

Baird, S.J.S., Cohen, J.T., Graham, J.D., Shlyakhter, A.I., Evans, J.S. (1996). Noncancer risk assessment: a probabilistic alternative to current practice. Human and Ecological Risk Assessment 2, 79-102.

Baird, S.J.S., Slob, W., Jarabek, A.M. (2001). Probabilistic noncancer risk estimation. Comments on Toxicology 7, 541-574.

Bokkers, B.G., Slob, W. (2005). A comparison of ratio distribution based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. Toxicological Sciences 85, 1033-1040.

Bokkers, B.G.H., Mengelers, M.J., Bakker, M.I., Chiu, W.A., Slob, W. (2017). APROBA-Plus: A probabilistic tool to evaluate and express uncertainty in hazard characterization and exposure assessment of substances. Food and Chemical Toxicology 110, 408-417.

Bosgra, S., Bos, P.M.J., Vermeire, T.G., Luit, R.J., Slob, W. (2005). Probabilistic risk characterization: an example with di(2-ethylhexyl)phthalate. Regulatory Toxicology and Pharmacology 43, 104-113.

Chiu, W.A., Axelrad, D.A., Dalaijamts, C., Dockins, C., Shao, K., Shapiro, A.J., Paoli, G. (2018). Beyond the RfD: Broad application of a probabilistic approach to improve chemical dose-response assessments for noncancer effects. Environmental Health Perspectives 126, 067009.

Chiu, W.A., Slob, W. (2015). A unified probabilistic framework for dose-response assessment of human health effects. Environmental Health Perspectives 123, 1241-1254.

Cho, W.-S., Han, B.S., Nam, K.T., Park, K., Choi, M., Kim, S.H., Jeong, J., Jang, D.D. (2008). Carcinogenicity study of 3-monochloropropane-1,2-diol in Sprague-Dawley rats. Food and Chemical Toxicology 46, 3172-3177.

Cooke, R. (2010). Conundrums with uncertainty factors Risk Analysis 30, 330-339.

Crump, K.S., Chiu, W.A., Subramaniam, R.P. (2010). Issues in using human variability distributions to estimate low-dose risk. Environmental Health Perspectives 118, 387-393.

Davison, A.C., Hinkley, D.V. (2009). Bootstrap methods and their application, Cambridge University Press, Cambridge.

Dorne, J.L.C.M., Walton, K., Renwick, A.G. (2005). Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review. Food and Chemical Toxicology 43, 203-216.

ECHA, European Chemicals Agency. (2012). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1, November 2012, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

Efron, B., Tibshirani, R. (1994). An introduction to the bootstrap, Chapman & Hall, Boca Raton.

Fahrmeier, L., Künstler, R., Pigeot, I., Tutz, G. (2001). Statistik. Der Weg zur Datenanalyse, Springer-Verlag, Berlin Heidelberg New York.

Gelman, A., Carlin, J.B., S., S.H., Rubin, D.R. (1998). Bayesian data analysis, Chapman & Hall, London.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). Markov chain Monte Carlo in practice Chapman & Hall, London.

Goble, R., Hattis, D. (2010). Are conundrums with uncertainty factors an obstacle to developing probabilistic interpretations of noncancer risks from chemicals? Risk Analysis 30, 340-345; author reply 353.

Hattis, D., Baird, S., Goble, R. (2002). A Straw Man Proposal for a Quantitative Definition of the RfD, in Final Technical Report, U.S. Environmental Protection Agency STAR grant # R825360, Human Variability in Parameters Potentially Related to Susceptibility for Noncancer Risks. Full version available at http://www2.clarku.edu/faculty/dhattis ; shortened version in Drug and Chemical Toxicology 2002; 25:403-436.

Hattis, D., Lynch, M.K. (2007). Empirically Observed Distributions of Pharmacokinetic and Pharmacodynamic Variability in Humans Implications for the Derivation of Single Point Component Uncertainty Factors Providing Equivalent Protection as Existing RfDs. In: J.C. Lipscomb, E.V. Ohanian (Eds), Toxicokinetics in risk assessment, pp. 69-93.

Lambert, B. (2018). A student's guide to Bayesian statistics, SAGE publications, London.

Mekel, O., Fehr, R. (2014). Probabilistic methods in toxicology. In: F.-X. Reichl, M. Schwenk (Eds), Regulatory Toxicology, Berlin, Heidelberg, pp. 253-259.

Monteith, K., Carroll, J., Seppi, K., Martinez, T. (2011). Turning Bayesian Model Averaging into Bayesian Model Combination. Proceedings of the International Joint Conference on Neural Networks IJCNN'11, 2657-2663.

NRC, National Research Council. (2009). Science and Decisions: Advancing Risk Assessment, Washington, DC: The National Academies Press. https://www.nap.edu/catalog/12209/science-and-decisions-advancing-risk-assessment.

Oldenkamp, R., Huijbregts, M.A., Ragas, A.M. (2016). Uncertainty and variability in human exposure limits - a chemical-specific approach for ciprofloxacin and methotrexate. Critical Reviews in Toxicology 46, 261-278.

Price, P.S., Keenan, R.E., Swartout, J.C., Gillis, C.A., Carlson-Lynch, H., Dourson, M.L. (1997). An approach for modeling noncancer dose responses with an emphasis on uncertainty. Risk Analysis 17, 427-437.

Renwick, A.G., Lazarus, N.R. (1998). Human variability and noncancer risk assessment - An analysis of the default uncertainty factor. Regulatory Toxicology and Pharmacology 27, 3-20.

Schneider, K., Hassauer, M., Oltmanns, J., Schuhmacher-Wolz, U., Elmshäuser, E., Mosbach-Schulz, O. (2004). Wahrscheinlichkeitsrechnung als Hilfsmittel zu Wirkungsabschätzung bei Arbeitnehmern. Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Fb 1012, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund/Berlin/Dresden, Wirtschaftsverlag NW Bremerhaven.

Schneider, K., Schuhmacher-Wolz, U., Hassauer, M., Darschnik, S., Elmshäuser, E., Mosbach-Schulz, O. (2006). A probabilistic effect assessment model for hazardous substances at the workplace. Regulatory Toxicology and Pharmacology 44, 172-181.

Seynaeve, D., Verbeke, T. (2017). Software for Monte Carlo simulation of a simple risk assessment. EFSA supporting publication 2017:EN-1316, doi:10.2903/sp.efsa.2017.EN-1316. http://www.efsa.europa.eu/de/supporting/pub/en-1316.

Shao, K., Shapiro, A.J. (2018). A Web-Based System for Bayesian Benchmark Dose Estimation. Environmental Health Perspectives 126, 017002.

Simon, T.W., Zhu, Y., Dourson, M.L., Beck, N.B. (2016). Bayesian methods for uncertainty factor application for derivation of reference values. Regulatory Toxicology and Pharmacology 80, 9-24.

Slob, W., Bakker, M.I., te Biesebeek, J.D., Bokkers, B.G.H. (2014). Exploring the uncertainties in cancer risk assessment using the Integrated Probabilistic Risk Assessment (IPRA) approach. Risk Analysis 34, 1401-1422.

Slob, W., Pieters, M.N. (1998). A probalistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. Risk Analysis 18, 787-798.

Swartout, J.C., Price, P.S., Dourson, M.L., Keenan, R.E. (1998). A probabilistic framework for the reference dose. Risk Analysis 18, 271-282.

van der Voet, H., Slob, W. (2007). Integration of probabilistic exposure assessment and probabilistic hazard characterization. Risk Analysis 27, 351-371.

van der Voet, H., van der Heijden, G.W., Bos, P.M., Bosgra, S., Boon, P.E., Muri, S.D., Brüschweiler, B.J. (2009). A model for probabilistic health impact assessment of exposure to food chemicals. Food and Chemical Toxicology 47, 2926-2940.

Vermeire, T., Stevenson, H., Pieters, M.N., Rennen, M., Slob, W., Hakkert, B.C. (1999). Assessment factors for human health risk assessment: a discussion paper. Critical Reviews in Toxicology 29, 439-490.

Vernez, D., Fraize-Frontier, S., Vincent, R., Binet, S., Rousselle, C. (2018). Experiencing a probabilistic approach to clarify and disclose uncertainties when setting occupational exposure limits. International Journal of Occupational Medicine and Environmental Health 31, 475-489.

Vose, D. (1996). Quantitative Risk Analysis: A guide to Monte Carlo simulation modelling, John Wiley & Sons, Chichester.

WHO, World Health Organization. (2014). Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. IPCS Harmonization Project Document 11, WHO, Geneva, Switzerland. http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf.

List of Tables

List of Figures

Annex 1: Examples

Both examples are based on oral studies. In order to keep these examples simple, no conversion to air concentrations is performed. As at this stage of the project no conclusions on adequate distributions for assessment factors can be drawn, the distributions as proposed by WHO in the APROBA tool are used. Although these distributions are meant for assessing risks in the general population, they are useful for demonstrating how the tools work and results can be interpreted. But note that using these distributions by no means does imply that they are recommended for deriving OELs.

Quantal data set

Input data – dose response data

Example dataset for quantal data:
- Substance: 3-monochloropropane-1,2-diol
- Study type: chronic toxicity study in rats
- Effect: renal tubule hyperplasia in male rats
- Source: Cho et al. (2008)

Dose-response data used for analysis:

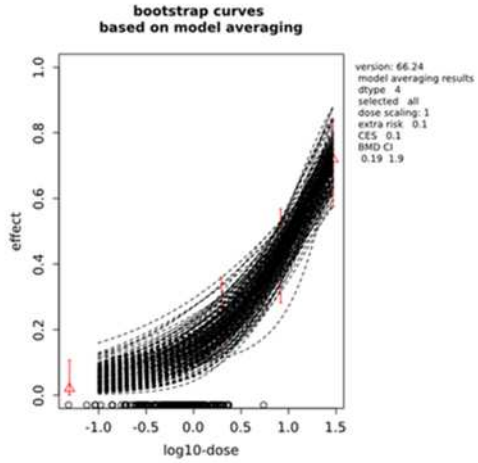
dose (mg/kg bw/day	effect # affected animals	n # animals in group
0.00	1	50
1.97	11	50
8.27	21	50
29.50	36	50

Dose-response modelling with the EFSA Benchmark dose modelling tool,

- with BMR = 10% extra risk compared to the controls
- with model averaging

yielded the following result:

BMDL	BMDU
0.19 mg/kg bw/day	1.88 mg/kg bw/day



Input data – uncertainty distributions

Allometric Scaling, based on rat weight 0.4 kg and human weight 70 kg:

- Median of allometric exponent: 0.7
- Standard deviation of estimate of exponent using in allometric scaling by body weight (based on 95% CI of 0.66 - 0.74)
- results in lognormal distribution of allometric scaling factor for rats (body weight 400 g) versus humans (70 kg):
    - 5 percentile 3.83
    - 95 percentile 5.97

Remaining interspecies variability:

- Median estimate of remaining chemical-specific TK and TD uncertainty after allometric scaling: 1
- Geometric standard deviation (GSD) of chemical-specific TK and TD uncertainty after allometric scaling: 1.95
- results in distribution with:
    - 5 percentile 0.33
    - 95 percentile 3

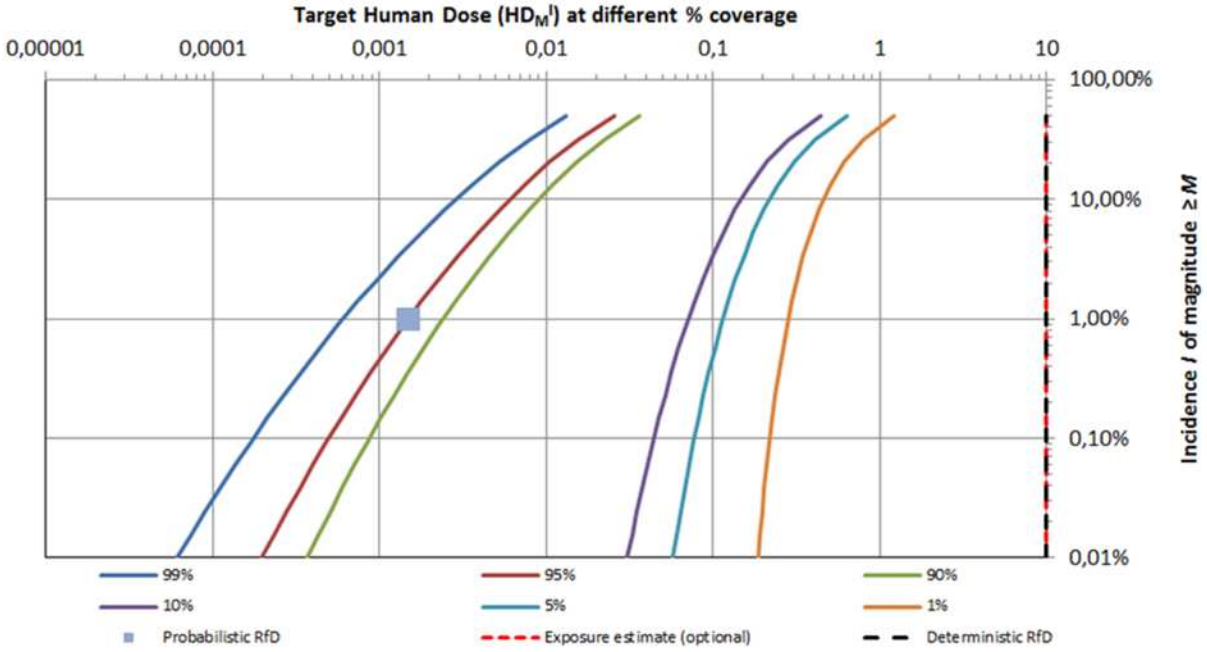
Intraspecies variability (humans):

- Median estimate of the Log(GSD_H) for human variability: 0.324
- GSD_U of the Log(GSD_H) for human variability = (P95/P50)^(1/1.645): 1.59
- results in (approximated) distribution with:
    - 50 percentile 9.69
    - 95 percentile 41.88

Modelling results

APROBA EXCEL®

Probabilistic guidance value =	= Approximate probabilistic HD_M^I at specified % confidence
0.0011 mg/kg bw/day	= Estimate of dose (mg/kg body weight per day) at which, with **95%** confidence of the population will **1%** have **renal hyperplasia of magnitude** ≥ **10%**



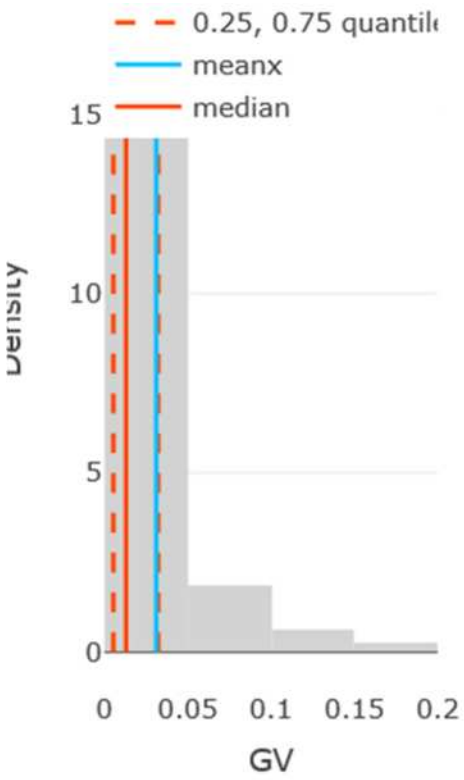
UNCERTAINTY ANALYSES	**% contribution**
ASPECT	**to overall uncertainty**
PoD	28%
Allometric scaling	1%
Interspecies TK/TD	26%
Intraspecies	46%
Greatest contributor to overall uncertainty	**Intraspecies**

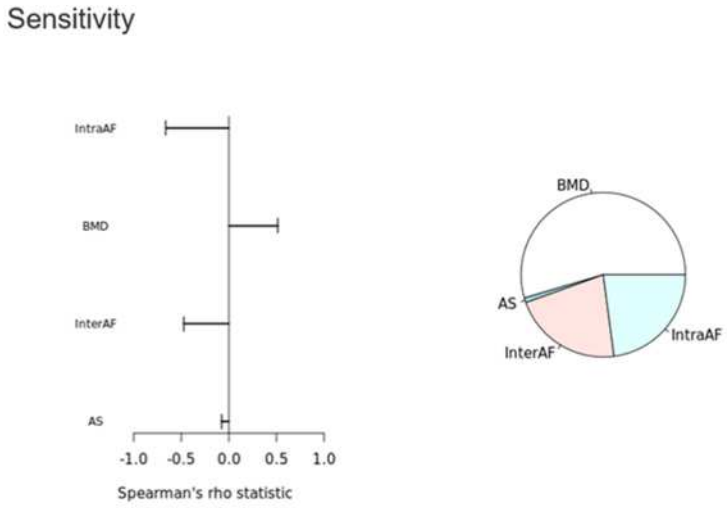
EFSA tool

Summary

variable	mean	sd	1%	5%	50%	90%	95%	99%
BMD	0.7606	0.6049	0.1188	0.1904	0.5931	1.441	1.876	3.049
AS	4.741	0.5997	3.514	3.829	4.699	5.533	5.79	6.359
InterAF	1.246	0.9456	0.2131	0.3322	0.9915	2.361	2.996	4.639
IntraAF	14.31	15.72	1.212	2.195	9.712	30.14	41.48	77.93
GV	0.03121	0.06231	0.0006273	**0.00154**	0.01311	0.07234	0.1133	0.2874

The 5th percentile of the distribution of the guidance value is 0.0015 mg/kg bw/day.





Comparison and discussion

The full probabilistic assessment carried out with the EFSA tool leads to a value with a 95% probability (5th percentile of the distribution) of 0.00154 mg/kg bw/day, compared to APROBA, which results in a value of 0.0011.

POD / APROBA HD_M^I = factor 173

POD / EFSA GV = factor 123.

It is reasonable that the full probabilistic model is somewhat less conservative compared to the approximated model (WHO 2014). The results are reasonably close.

Note that the absolute figures and the absolute distance between POD and the obtained values are not meaningful, as long no distributions adequate for worker assessments are used.

The sensitivity analysis shows differences: Intraspecies extrapolation is considered to contribute most to uncertainty in APROBA, whereas the BMD accounts for the highest uncertainty in the EFSA tool. The reason for this discrepancy is that different algorithms are used for the sensitivity analysis: in APROBA for each variable the spread of the distribution is calculated relative to the sum of spreads, irrespective of whether it is placed in the enumerator or denominator. In the EFSA tool the impact of each variable on the outcome parameter is calculated on a normal scale. This implies that variables in the enumerator (BMD) get a higher sensitivity score than those in the denominator (José Cortinas Abrahantes, EFSA, personal communication, 18 Sept 2019).

Continuous dataset

Input data – dose response data

Example dataset for quantal data:
- Substance: nalidixic acid
- Study type: chronic toxicity study in rats
- Effect: body weight changes in male and female rats
- Source: NTP (1989), NTP TR No. 368

Dose-response data used for analysis:

	concentration in food (ppm)	bw	SEM	n	sex
7	0	29.4	1.13	10	f
8	1000	28.2	1.18	10	f
9	2000	28.7	1.06	9	f
10	4000	27.1	0.42	10	f
11	8000	24.8	0.84	10	f
12	16000	23.6	0.55	10	f
1	0	36.1	0.89	10	m
2	1000	35.0	0.64	10	m
3	2000	34.9	0.71	10	m
4	4000	33.6	0.41	10	m
5	8000	32.4	0.47	10	m
6	16000	31.4	0.71	10	m

Dose-response modelling with the EFSA Benchmark dose modelling tool,

- with BMR = 10% relative difference in final body weight compared to the controls
- with model averaging

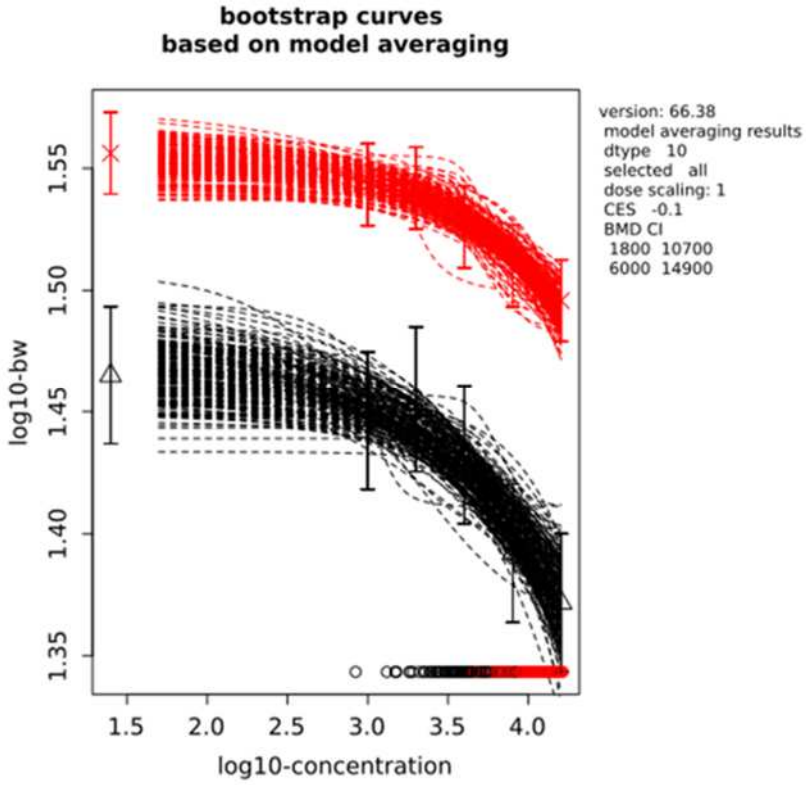
yielded the following result:

endpoint	subgroup	BMDL (ppm food)	BMDU (ppm food)
bw	f	1800	10700
bw	m	5950	14900

For the more sensitive female rats these food concentrations can be converted to a dose (with defaults as given in ECHA (2012)):

1800 ppm, with a food factor of 50 g per kg bw per day: 90 mg/kg bw/day
10700 ppm, with a food factor of 50 g per kg bw per day: 535 mg/kg bw/day

Note: in the EFSA tool with 90 and 535 mg/kg bw/day as 5th and 95th percentile, no lognormal distribution could be fitted. Instead a Weibull function with a similar shape was used.



Input data – uncertainty distributions

See previous chapter, apart from:

Allometric Scaling, for female rat weight 0.35 kg and human weight 70 kg:

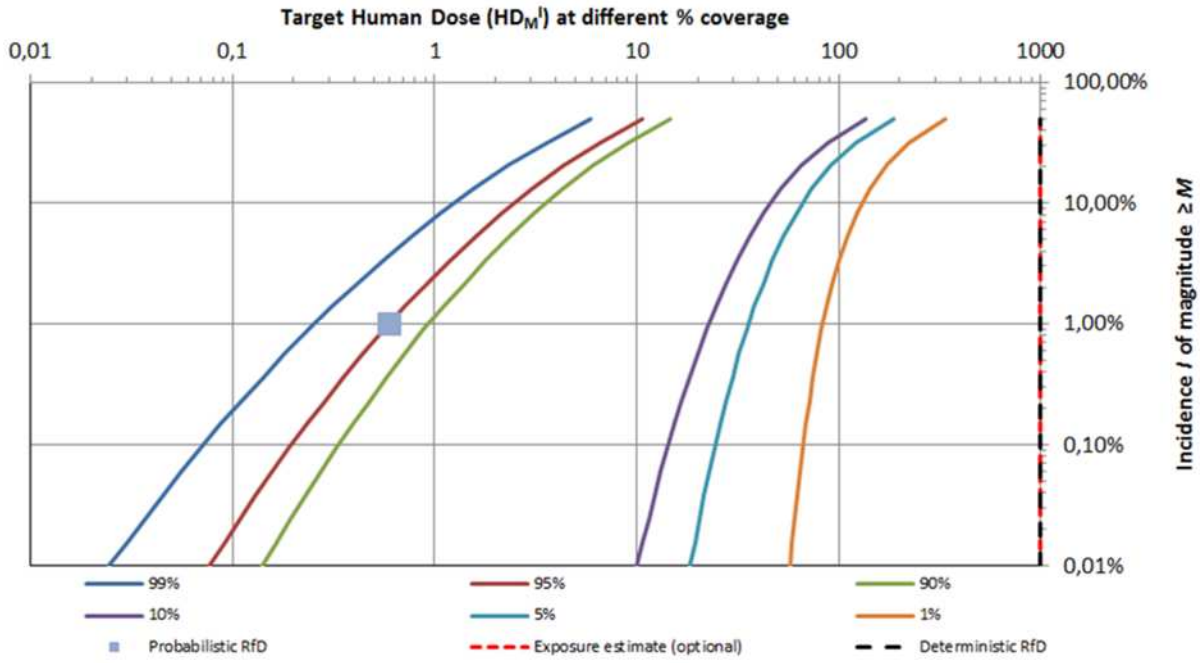
- Median of allometric exponent: 0.7
- Standard deviation of estimate of exponent using in allometric scaling by body weight (based on 95% CI of 0.66-0.74)
- results in normal distribution of allometric scaling factor for female rats (body weight 350 g) versus humans (70 kg):
    - 5 percentile 3.97
    - 95 percentile 6.06

Note: in the EFSA tool with these percentile values, no lognormal distribution could be fitted. Instead, a normal function with a similar shape was used.

Modelling results

APROBA

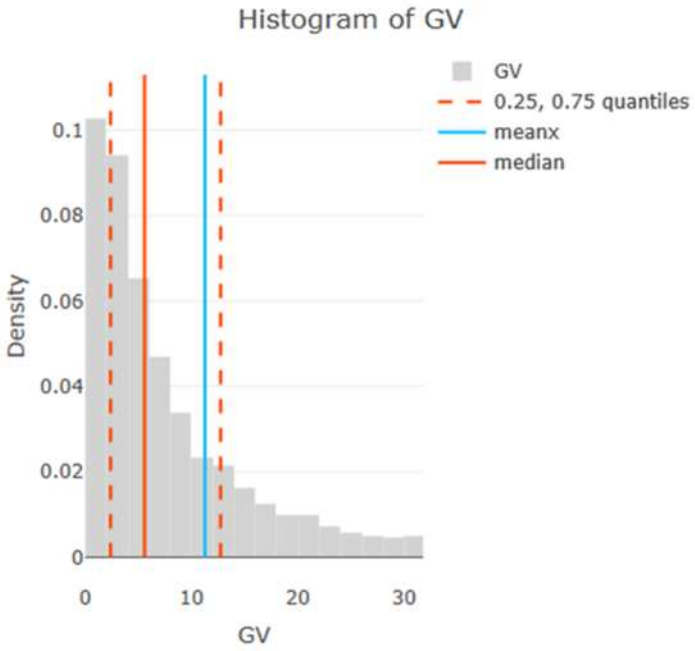
Probabilistic guidance value =	= Approximate probabilistic HD_M^I at specified % confidence		
0.597	= Estimate of dose (mg/kg body weight per day) at which, with		
	95%	confidence	
	1%	of the population will have	**reduced body weight development**
	of magnitude	**≥**	**10%**



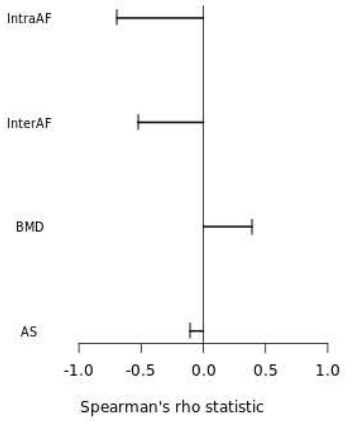
ASPECT	% contribution to overall uncertainty
PoD	19%
Interspecies scaling	1%
Interspecies TK/TD	29%
Intraspecies	51%
Greatest contributor to overall uncertainty	**Intraspecies**

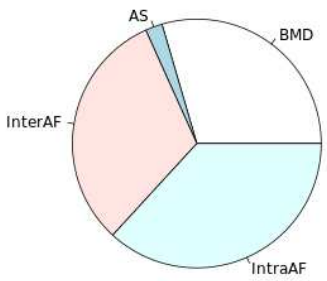
EFSA tool

variable	mean	sd	1%	5%	10%	50%	90%	95%	99%
BMD	294.1	135.3	43.91	91.78	126.8	283.6	475.5	537.4	645.1
AS	5.016	0.6365	3.524	3.973	4.196	5.012	5.822	6.07	6.5
InterAF	1.243	0.9726	0.2108	0.3292	0.4178	0.9824	2.362	3.017	4.85
IntraAF	14.33	15.5	1.248	2.26	3.102	9.63	30.55	41.45	75.72
GV	11.27	18.45	0.2754	**0.6649**	1.096	5.587	26.16	39.64	87.83



Sensitivity





Comparison and discussion

The full probabilistic assessment carried out with the EFSA tool leads to a very similar value as APROBA: with 95% probability (5^{th} percentile of the distribution) GV is 0.66 mg/kg bw/day, compared to the value of 0.60 mg/kg bw/day obtained with APROBA.

POD / APROBA HD_M^I = factor 151

POD / EFSA GV = factor 136.

Note that the absolute figures and the absolute distance between POD and the obtained values are not meaningful, as long no distributions adequate for worker assessments are used.

The sensitivity analysis again shows some differences (see above for explanations): intraspecies extrapolation is contributing most to uncertainty in both tools, but in the EFSA tool uncertainty is more even distributed between intraspecies extrapolation, POD and interspecies extrapolation.

Annex 2: Explanations

Distributions

Empirical distribution (frequency distribution)

If a parameter x is measured many times, the more likely values will be measured more often. A graphical presentation is given in the next figure. The frequency distribution of measured continuous parameters can be visualised as such a histogram (Fahrmeier et al. 2001). A histogram approximates the underlying theoretical probability density function by a step function. It is generated by splitting the range of measured values into intervals and by showing for each interval the number of values in this interval.[2]

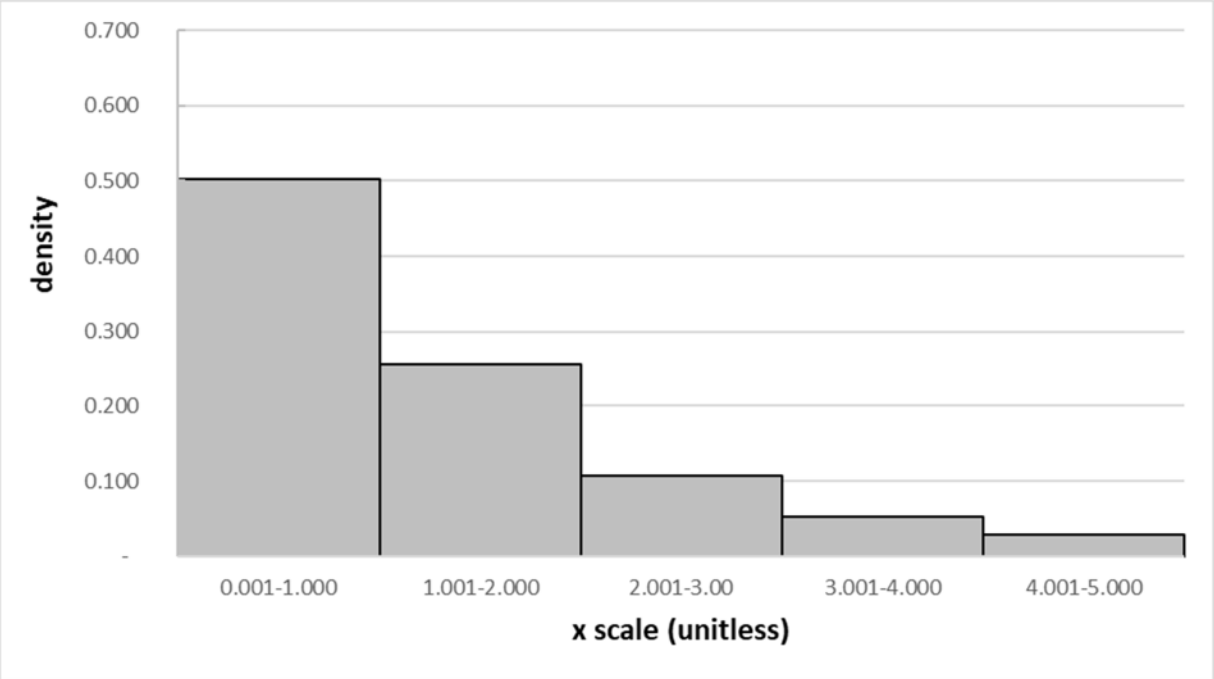


Figure A-1 Example for an empirical distribution (x axis unitless)

[2] In a normalised histogram the x axis gives the proportion (= number of values in the interval / number of all values) of values in this interval as a rectangle over the interval which has an area equal to this proportion. The area is calculated as interval width x rectangle height, and the rectangle height is chosen such that the sum over all rectangle areas is equal to 1. Usually, all intervals have the same width, though this is not a necessary requirement. The number of intervals together with the interval limits control the appearance of the histogram. For equidistant interval limits there are recommendations regarding the number of intervals to employ, which aim at providing a good representation of the distribution shape.

Model fitting

Curve fitting may lead to a parametric model representing the empirical distribution with high accuracy. Such curve fitting in the case of the empirical dataset above might lead to a lognormal distribution by its expected value μ and its standard deviation σ, both expressed on the log scale. The parameter μ is estimated from data as the arithmetic mean of the logarithms of the data), and the parameter σ is estimated by the standard deviation of the data logarithms (natural logarithms or logarithm to base 10 can be used).

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp(-\frac{(\ln(x) - \mu)^2}{2\sigma^2})$$

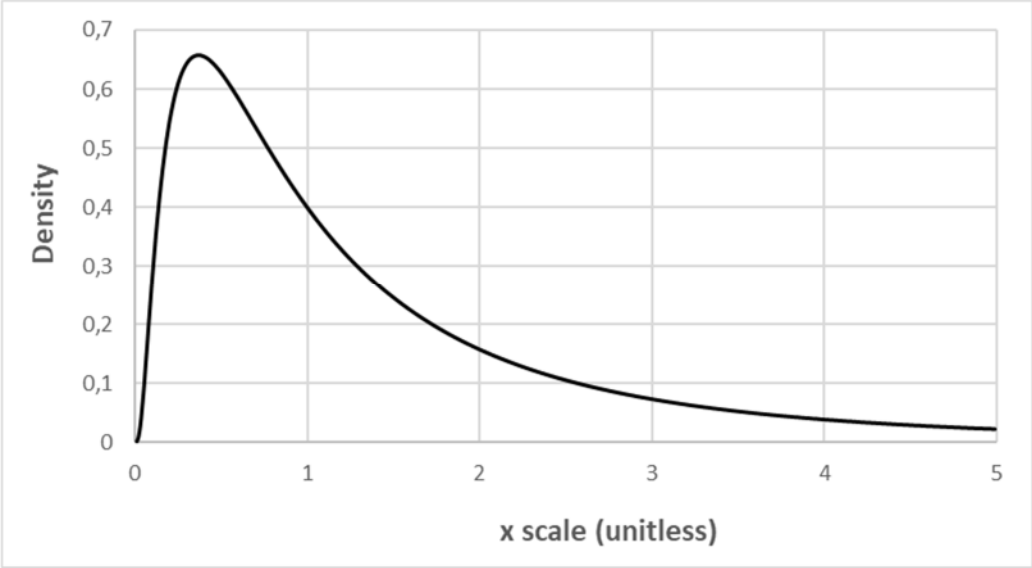
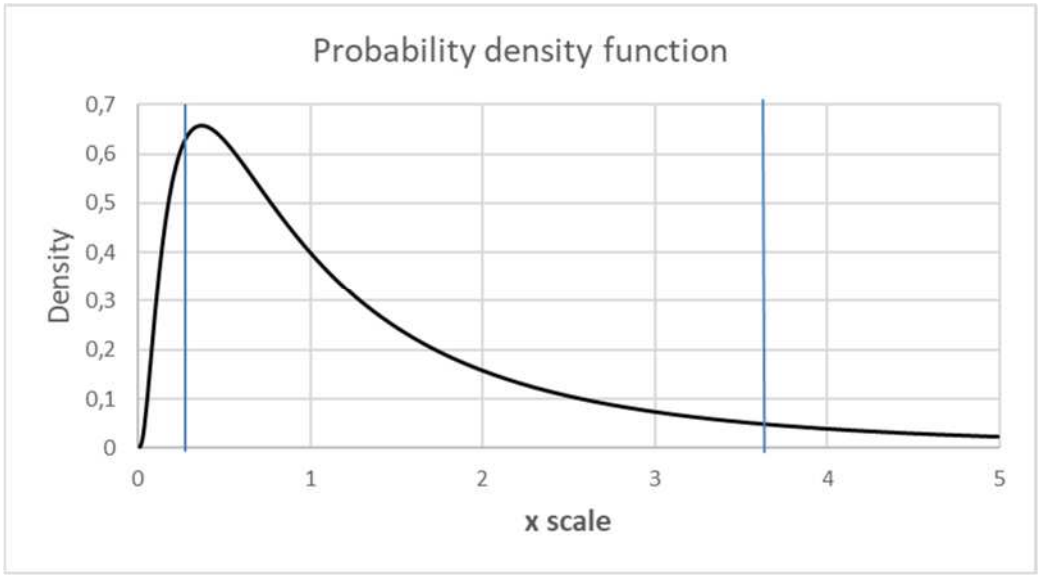


Figure A-2 Probability density of a log-normal distribution

Probability density function versus cumulative distribution function

The distribution above shows a probability density function (PDF). Its advantage is to show the shape of the distribution more clearly than the alternative presentation as cumulative distribution function (CDF) discussed below. The probability of getting values lying between two limits x_1 and x_2 is the area below the curve (integral) between x_1 and x_2. The 10th percentile is the point on the x axis, under which 10% of all values lie (and, correspondingly, 90% are larger than the 10th percentile). Blue lines in the figure below show the 10th and 90th percentile for the density above.

An alternative presentation of a distribution is the cumulative distribution function, which has values between 0 and 1. For each x value the y axis gives the probability that the parameter has a value of x or below. The 10th and 90th percentiles are the values on the x-axis with y values of 0.1 and 0.9, resp. (blue lines in figure below). The advantage of a CDF is that probabilities can be read directly from the graph.



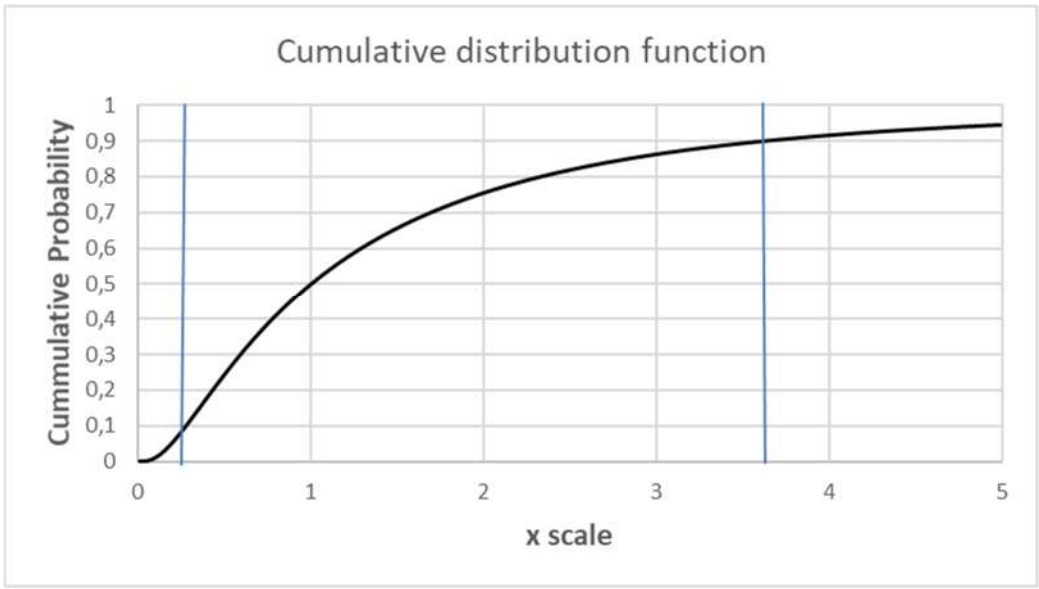


Figure A-3 Probability density versus cumulative distribution function (with 10th and 90th percentiles as blue lines)

Explanation of terms

Bayesian methods

Bayesian methods is the summary term for a certain philosophy in statistics. It starts from the assumption that there is certain prior information about the problem to solve, typically from earlier studies. The prior information is typically an assumption about the distribution of the quantity to analyse. Prior information and data are combined using Bayes' law of total probability and result in a posterior distribution, which is the central

result of the analysis. As an example, highest posterior density regions are used as equivalent for confidence intervals known from standard (frequentist) analysis (Gelman et al. 1998; Lambert 2018).

The concept of Bayes analysis is old but has for a long time come out of focus, because the computational effort was too large. Meanwhile, due to fast computers and some mathematical developments, Bayesian approaches have become feasible. The philosophical counterpart to Bayesian analysis is frequentist analysis, which does not use prior information, but derives its results only from the analysed data.

Monte-Carlo simulation

Monte Carlo simulation (also MC experiment or MC analysis) is a mathematical approach in statistics for deriving (among others) the distribution of an estimated quantity or of a test statistic. It is also used to analyse the properties of a sampling design, of an experimental design or of more general stochastic processes. It is typically applied, when it is difficult or impossible to derive this distribution analytically. Finding the distribution of an output statistic that results from combining values from various input distributions by a mathematical formula is an example for such a problem.

The idea of MC simulation is to generate many fictive input data sets of the type under study. From each data set, the interesting statistic (estimated value or test statistic) is calculated. Doing this for each fictive data set generates the desired distribution of the statistic. Often this distribution itself is of interest. It can also be used to e.g. calculate a confidence interval for an estimated parameter or the p value for an observed test statistic.

The basic operation in an MC simulation is generating the fictive input data sets. The procedure can be illustrated by using **Figure 1-1**. There, four input variables (POD, AF1, AF2, AF3) are combined to give the output GV by the equation GV = POD/(AF1 * AF2 * AF3). All input variables are random variables. The distribution of each input variable is completely known (i.e. including the numerical values of the distribution parameters like mean value or standard deviation). Each input may have its own distribution. Together, all input distributions induce a distribution of GV, which is the quantity of interest. In this example, the formal solution requires solving a high-dimensional integral, which is not possible.

For the MC solution, *n* fictive datasets are generated. Each dataset consists of four numbers, one value for each of the four input variables. Each number is a computer generated random number from the relevant distribution. Such a random number is calculated by first drawing a number from a uniform distribution. All possible values of a uniform distribution have the same probability like (theoretically) the possible outcomes when playing roulette, hence the name “Monte Carlo” method. The uniform random number is then transformed to a random number having the required input distribution. The four transformed numbers are then combined by the above equation to obtain (one realisation of) the output quantity GV. Repeating the process of data

generation and calculation of GV *n* times gives the desired distribution of GV. The larger *n* is, the better is the knowledge of the GV distribution.

This description of an MC simulation for independent inputs refers to the simplest form of a MC simulation. However, it already makes obvious that the MC simulation provides a result only for exact those input conditions that were used to generate the fictive data sets. Also the number *n* of data sets is not obvious to determine. A large *n* is desirable, but can generate large computing time. Therefore, methods have been developed to accelerate random number generation. Special action is also required, when input variables are not statistically independent from another, as assumed in the example.

MC analysis has the advantage of being conceptually simple and of being nearly universal in the sense that many problems can be solved that cannot be solved by other methods. MC analysis has the disadvantage of being (computer-) time-consuming, with some uncertainty whether the number *n* of generated datasets is sufficient. Also, different from a formal solution, MC analysis gives no structural insights into the problem, but provides only a numerical solution under exactly the conditions that were used in the simulation. If a parameter of the input distribution is changed, the whole MC calculation must be done again, while the formal solution, if it exists, provides an answer for each input distribution parameter without new derivation (Gilks et al. 1996; Monteith et al. 2011; Vose 1996).

Random sampling, Latin-Hypercube sampling, Markov Chain Monte Carlo sampling

There are several possibilities to perform drawings:

Random sampling means drawing values randomly from a distribution, without further conditions on the sampling process. The distribution is either given by a formula or by a density estimate or as the empirical distribution of a data set. The “random” component in this operation means that in the initial step of random number generation, described in the previous section, uniformly distributed numbers from the whole range [0, 1] are used. As in practice the number of values is always finite, the distribution of the numbers actually drawn will not exactly be uniform. This causes transformed random numbers which do not have exactly the theoretically required form.

A random sample has the advantage that its statistical properties can easily be generalized to the underlying universe. However, random sampling requires large samples to ensure that generated numbers have the intended distribution. This holds especially if properties of an extreme part of the distribution are sought (e.g. the location of the 0.1% quantile). This has led to the development of other sampling schemes as described below.

Stratified sampling is an enhancement of random sampling, which subdivides the range of possible input values into disjoint subgroups (strata), then first selects a stratum to sample from and subsequently takes a random sample from the selected stratum. In this way, a good coverage of the universe can be achieved with smaller

samples than under random sampling. This approach is also applicable for sampling multidimensional input variables.

Latin hypercube sampling is an enhancement of stratified sampling. It is particularly useful for sampling multidimensional input variables, when not all input variables are of same importance to the outcome. Latin hypercube sampling ensures that the full range of each input variable gets represented by the sample. If certain conditions hold for the relationship between input and outcome variables, then the variance of the output variable under Latin hypercube sampling is less than or equal to the variance under random sampling.

Markov Chain Monte Carlo (MCMC) is a concept to generate vectors of dependent random numbers. These are needed e.g. for an MC analysis, which deals with problems involving statistically dependent input random variables. In the example from the previous chapter such a dependency would occur if the probability of AF1 having a certain value would depend on the value of POD. In this case, random values for AF1 and POD can no more be obtained by independent drawing of these two values from the distributions of AF1 and POD, respectively. Instead, the pair (AF1, POD) must be drawn from the joint distribution of AF1 and POD. The joint distribution is not the product of the two single distributions, but must be obtained from corresponding two-dimensional data on the joint occurrence of AF1 and POD. A simple example of a two-dimensional distribution is that ln(AF1) and ln(POD) have a joint normal distribution, which is characterized by a vector of means (μ_{AF1}, μ_{POD}) and a 2 x 2 covariance matrix containing the variances $(\sigma_{AF1})^2$ and $(\sigma_{POD})^2$ and the covariance $\sigma_{AF1,POD}$, which quantifies the degree of dependency between both variables.

The MCMC method starts like a usual MC analysis, but then generates vectors of dependent input variables by feeding independent random numbers into a Markov chain. A **Markov chain** is a stochastic process in discrete space and over discrete time, where the state of the chain at time t+1 depends only on the state at time t, but not on earlier states. This property allows a simple computer generation of dependent random numbers (Gelman et al. 1998; Gilks et al. 1996).

Bootstrapping

Bootstrapping (BS) is a technique to obtain properties of a statistic derived from data by resampling new data sets from the same data from which the statistic was derived. It is closely related to MC simulation, because the desired properties result from computing the statistic from randomly selected data sets. However, BS samples its data sets only on basis of the real data, not from distributions that were selected from additional considerations. There are several versions of bootstrapping: empirical BS takes random samples from the real data set, parametric BS fits a parametric distribution to the real data sets and generates random numbers from the fitted distribution. Semi-parametric BS fits a nonparametric density to the data (e.g. by a kernel density estimate) and gets random numbers from this density (Davison and Hinkley 2009; Efron and Tibshirani 1994).

REPORT 5:
Route-to-Route Extrapolation

RESEARCH PROJECT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Summary

In regulatory toxicology route-to-route extrapolation can be applied if reliable route-specific data is lacking for the uptake pathway of interest. In the absence of inhalation studies route-to-route extrapolation (from oral to inhalation exposure) may be considered for deriving OELs or OEL-analogue values.

Specific criteria and conditions have been proposed for applying route-to-route extrapolation by the Interdepartmental Group on Health Risks from Chemicals (IGHRC) and Geraets et al. (2014). Further, the ECHA Guidance document on Information Requirements and Chemical Safety Assessment, R.8 provides practical support for performing the individual extrapolation steps, starting from a route-specific point of departure.

Other institutions like SCOEL or ECETOC base their suggested procedure for route-to-route extrapolation mainly on the ECHA R8 guidance. Similarly, the German “Ausschuss für Gefahrstoffe” (AGS) advises against route-to-route extrapolation if critical differences between routes exist, e.g., in case of a relevant first-pass effect. In the absence of other evidence, it is assumed that the amount of substance administered orally has the same efficacy as the amount inhaled.

There is a general agreement that route-to-route agreement is applicable only, if the expected critical effects are of systemic, not local nature and if no differences exist that make predictions for the other route unreliable, for example a severe first-pass effect.

1 Introduction

Route-to-route extrapolation is used in regulatory toxicology in case reliable route-specific data is lacking for the uptake pathway of interest. Toxicity testing is mostly performed with oral administration; inhalation and dermal toxicity data are often absent. In such cases the use of route-to-route extrapolation (from oral to inhalation exposure) may be considered for deriving OELs or OEL-analogue values at the workplace.

The first systematic approaches go back to (Stokinger and Woodward 1958). They performed route-to-route extrapolation of guideline values and derived drinking water values based on existing occupational exposure limits. Route-to-route extrapolations are applied by various institutions and are defined as the prediction of an equivalent dose for the route of interest that produces the same response.

The very basic condition for route-to-route extrapolation is that comparable toxicological effects are expected for the considered exposure pathways. Ideally, kinetics, metabolism, and toxicity of a systemically active substance should be comparable across the different pathways, or the differences should be known or predictable.

For the application of route-to-route extrapolation it is necessary to have a common measure that transforms the external exposure (e.g., oral dose) to an internal exposure. PBPK models (physiologically based pharmacokinetic models) can provide information on route-specific target tissue concentrations (see section 0. for details). In the absence of such a model however, a route-specific absorbed dose can be estimated based on the external dose (in mg/kg bw) while considering the route-specific absorption fraction.

By using this method, it is possible to transform doses obtained for an exposure route (e.g., No Observed Adverse Effect Level (NOAEL)) to other routes of exposure.

2 Route-to route extrapolation in regulatory contexts

2.1 Current practice for derivation of OEL values

In Report 1 of the current project (Comparison of methods for deriving OELs and OEL-analogue values, Table 2-4) information on route-to-route extrapolation in the different guidance documents for derivations of OEL values are provided. For a comparison this table can be consulted.

In the „Bekanntmachung zu Gefahrstoffen (BekGS 901)“ the German AGS published its position to route-to-route extrapolation for the derivation of OEL values (AGS 2010). AGS states that for systemic effects route-to-route extrapolation (from oral to inhalation) is possible if there is no evidence of significant differences in absorption and metabolism (e.g., first-pass effects). In the absence of such evidence, it is assumed that the amount of substance administered orally has the same efficacy as the amount inhaled. In addition, the different concentration-time patterns that may occur with gavage studies compared to feeding studies and the resulting differences in the toxic response are neglected. AGS further elaborates that route-to-route extrapolation is not possible

- if only repeated dermal studies are available, as there may be large differences in the amount of substance absorbed,
- in case of metal compounds, since there are indications in the literature of non-systematic and strongly varying absorptions after oral or inhalation exposure,
- for substances acting locally, since different organs may be affected and the effective doses presumably differ,
- for poorly soluble substances (solubility <1 mg/L H_2O; <1 mg/kg fat), since it is known that insoluble or poorly soluble substances may be non-toxic after oral exposure, but inhalation may lead to significant effects on the lungs (for these substances the general dust limit value must be applied),
- in case that only repeated oral studies are available combined with evidence for toxicologically significant dermal absorption.

The REACH guidance R.8 (ECHA 2012) and the “Guidelines on route-to-route extrapolation of toxicity data when assessing health risks of chemicals” published by (IGHRC 2006) propose default values for absorption in route-to-route extrapolation. These two guidelines are presented in more detail in the following chapter 0.

In the Plant Protection Products Directive (EC 2006) the derivation of AOELs for workers are usually based on NOAEL values from repeated oral toxicity studies. In case of route-to-route extrapolation from oral to dermal or inhalation exposure the actual oral absorption fractions instead of any defaults are used since this information

is part of the data requirements. For the dermal route, again route-specific absorption data should be used. If this is not available default values as assumed in the EFSA guidance on dermal absorption (EFSA 2012) are to be applied. In case of the inhalation route the absorption is set to 100% per default.

The ECHA Guidance on the Biocidal Products Regulation (ECHA 2017) points out that in general “route-to-route extrapolation is considered to be a poor substitute for toxicity data obtained using the appropriate route of exposure”. Several criteria are listed in the guidance that have to be met in order to perform route-to-route extrapolation (e.g. toxicity is a systemic effect not a local one). In case of the requirement of route-to-route extrapolation, information on the extent of absorption for the different routes of exposure should be used to modify the starting point. In the absence of relevant data for oral absorption or if available data does not indicate an absorption significantly below 100%, 100% should be assumed. For inhalation no defaults are provided.

With regard to absorption rates ECETOC (2010) cites an unpublished report by Verband der Chemischen Industrie from 2008. According to this report oral absorption for most organic substances is as high or higher than absorption by inhalation. Therefore, defaults of 100% for oral and inhalation absorption are used by AGS and MAK Commission. For metals, a low oral bioavailability is frequently observed, therefore, 50% oral absorption is assumed by MAK Commission as default (DFG 2018).

2.2 Guidelines on route-to-route extrapolation

In 2006 the Interdepartmental Group on Health Risks from Chemicals (IGHRC) published a guidance document “Guidelines on route-to-route extrapolation of toxicity data when assessing health risks of chemicals” explicitly dealing with route-to-route extrapolation of toxicity data. The document is “intended to provide general guidance to assist those undertaking toxicological risk assessments” (IGHRC 2006). In this document criteria are established that should be met to enable confidence in such extrapolations:

- Absorption is the same between routes, or the difference is known and can be quantified
- The critical target tissue is not at the portal of entry of the compound (i.e., the concern is with systemic toxicity and not local effects)
- There is no significant metabolism of the chemical by oral, gut or skin enzymes or in pulmonary macrophages, or transformation by other processes in the gut or lung;
- First-pass effects are minimal;
- The chemical is relatively soluble in body fluids.

Similar criteria were proposed by Geraets et al. (2014) for a reliable route-to-route extrapolation:

- The toxicological data are considered adequate and reliable
- The critical effect(s) for the routes of exposure under consideration are systemic, influence of local effects can be excluded
- The toxic effect is independent of the route of exposure
- The absorption is the same for the different routes of exposure OR the difference in absorption is known and can be quantified
- The half-life of the substance is long
- Hepatic first-pass metabolism is low
- No significant chemical transformation by intestinal microflora or pulmonary macrophages is expected
- The substance is considered relatively soluble in body fluids.

Due to the fact that most available experimental data are obtained using the oral route, the IGHRC document concentrates on extrapolation from the oral to the dermal or inhalation routes. IGHRC proposes the following absorption ratios in case of route-to-route extrapolation.

Oral → dermal: Ratio of 1 (in the absence of any data, with the assumption that dermal absorption will not exceed the oral absorption)

Oral → inhalation: Ratio of 2 (meaning that the absorption fraction for the oral route is half that of the inhalation route; applies for substances with high oral toxicity)

A ratio of 10 is proposed by IGHRC for substances with low oral toxicity,if a low toxicity potential in an oral study was observed and it is unknown whether the chemical has a high oral absorption and a low toxic potential or a low oral absorption and a high toxic potential.

In its considerations on route-to-route extrapolation the ECHA guidance R.8 (ECHA 2012) refers to the IGHRC document for specific criteria. There is a general consensus on the applicability of route-to-route extrapolation:

“If no adequate experimental effect data are available on the relevant route of exposure for the population under consideration, route-to-route extrapolation might be an alternative, however only for systemic effects, not for local effects (e.g. irritation of the lungs following inhalation of a substance). Even for systemic effects route-to-route extrapolation is considered appropriate only under certain conditions (e.g. no first pass effects)”. ECHA also points out that the use of physiologically based

pharmacokinetic (PBPK) modelling, would be of great value in route-to-route extrapolation (see section 0).

With regard to correction for differences in absorption in the absence of substance-specific data the ECHA guidance proposes to use a default factor of 2 in the case of oral-to inhalation extrapolation, which is equivalent to assume 50% absorption for the oral pathway and complete absorption by inhalation. For oral-to-dermal extrapolation the same absorption rate is assumed for both routes. These assumptions are in line with the proposals in the IGHRC document.

Further, in an appendix, the ECHA guidance provides detailed schemes to follow when deriving DNELs and starting from a POD of another exposure route.

Several other guidance documents refer to the ECHA guidance when it comes to route-to-route extrapolation (DFG 2019; ECETOC 2003; SCOEL 2017).

2.3 Recent evaluation

In a publication from 2016 uncertainties associated with route-to-route extrapolation were investigated by derivation of extrapolation factors based on no/lowest effect levels (NOELs/LOELs) from substances in the Fraunhofer RepDose® database (Schröder et al. 2016). For route-to-route extrapolation from oral to inhalation 246 study pairs on 110 substances were analysed. For systemic effect levels for inhalation studies derived from oral studies an extrapolation factor of 2.2 was obtained. An extrapolation factor of 3.2 was obtained when not distinguishing local or systemic effects. The authors recommend the use of a general factor of 3 for route-to-route extrapolation from oral to inhalation exposure in order to cover for the possibility that unexpected local effects may occur that trigger the LOEL.

For route-to-route extrapolation from oral to dermal 46 study pairs on 28 substances were analysed. An overall extrapolation factor of 0.4 was obtained for systemic effects. However, the authors highlight the limited number of analysed studies and consider their results as preliminary.

2.4 Route-to-Route extrapolation and PBPK modelling

A physiologically based pharmacokinetic (PBPK) model is the result of a mathematical modelling technique, which can predict toxicokinetic properties like absorption, distribution, metabolism and excretion (ADME) of substances.

Therefore, PBPK models give quantitative descriptions of toxicokinetic properties of chemicals in the body. PBPK models facilitate more scientifically sound extrapolations across studies, species, routes and dose levels (WHO 2010).

As pointed out by IGHRC (2006) and the ECHA guidance R.8 (2012) physiologically based pharmacokinetic (PBPK) modelling data are very useful for route-to-route extrapolation and would make the use of default values unnecessary. However, only

for a limited number of substances these data are available and cannot easily be generated by a risk assessor. To establish a reliable PBPK model ideally a combination of experimental studies should be available which put light on toxicokinetic properties, of the substance. The more data are available to inform the model, the more precise and reliable the PBPK model will be. With such a reliable model specified dose metrics of interest (e.g., absorbed dose after inhalation exposure) can be calculated.

Some publications are available describing the use of PBPK modelling for route-to-route extrapolation. For example (Gajewska et al. 2014) studied the application of PBPK modelling in oral-to-dermal extrapolation of cosmetic ingredients. They concluded that for a fixed external dose oral exposure does not always give higher internal concentrations than dermal exposure.

Sweeney and Gargas (2016) extrapolated oral data obtained for 1,2-dichlorethane in subchronic rat studies and an extended one-generation reproductive toxicity study (EOGRTS) using PBPK modelling. The authors found that the selection of the “internal metric” (e.g. plasma concentration of parent compound or total amount metabolised), which is used to establish route-to-route equivalency, influences the NOAEL-equivalent inhalation exposure concentration. Therefore, the authors conclude that the internal metrics are a key determinant of inhalation toxicity reference criteria. In the case of the studies selected by Sweeney and Gargas (2016) for 1,2-dichlorethane a factor of 17 was found between the results obtained with different input data (different oral studies for extrapolation in PBPK model and different assumptions or endpoints regarding metabolism).

3 Conclusion

In the absence of studies on inhalation exposure route-to-route extrapolation (from oral to inhalation) may be considered for deriving OELs or OEL-analogue values. There is a broad consensus that it should be applied only, if the expected critical effects are of systemic, not local nature and if no differences exist that make predictions for the other route unreliable, for example a severe first-pass effect.

Several guidance documents are available which can be consulted by a risk assessor when considering route-to-route extrapolation for an individual substance. The “Interdepartmental Group on Health Risks from Chemicals” (IGHRC 2006) as well as Geraets et al. (2014) proposed detailed criteria for applying route-to-route extrapolations.

In line with the IGHRC document the ECHA R.8 guidance (ECHA 2012) recommends to use a factor of 2 for extrapolating from oral to inhalation, if specific data are lacking. This is equivalent to assuming 50% (oral) and 100% (inhalation) absorption as defaults. For oral to dermal extrapolation no difference in absorption should be assumed, as a default. The ECHA guidance document further provides detailed and helpful guidance on the individual steps for deriving an inhalation guidance value, starting from oral data.

Also, the German (AGS 2010) advises to use route-to-route extrapolation for systemic effects if there is no evidence of significant differences in absorption and metabolism (e.g. first-pass effects). In the absence of other evidence, it is assumed that the amount of substance administered orally has the same efficacy as the amount inhaled.

If substance-specific data on absorption are available, these data should always be considered first and route-to-route extrapolation should be based on these data.

References

AGS, Ausschuss für Gefahrstoffe. (2010). Bekanntmachung zu Gefahrstoffen. Kriterien zur Ableitung von Arbeitsplatzgrenzwerten. BekGS 901. Das Gemeinsame Ministerialblatt 61 (32), 691-696. http://www.baua.de/cln_135/de/Themen-von-A-Z/Gefahrstoffe/TRGS/Bekanntmachung-901.html.

DFG, Deutsche Forschungsgemeinschaft. (2018). MAK- und BAT-Werte-Liste 2018. Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe. Mitteilung 54, WILEY-VCH Verlag GmbH, Weinheim. https://onlinelibrary.wiley.com/doi/book/10.1002/9783527818402.

DFG, Deutsche Forschungsgemeinschaft. (2019). MAK- und BAT-Werte-Liste 2019. Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe. Mitteilung 55, WILEY-VCH Verlag GmbH, Weinheim. https://onlinelibrary.wiley.com/doi/book/10.1002/9783527826155.

EC, European Commission. (2006). Guidance for the Setting and Application of Acceptable Operator Exposure Levels (AOELs). Draft. SANCO 7531 - rev.10, Directorate E – Safety of the food chain. E3 - Chemicals, Contaminants, Pesticides. https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_ppp_app-proc_guide_tox_accpt-exp-levs-2006.pdf.

ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals. (2003). Technical Report No. 86. Derivation of Assessment Factors for Human Health Risk Assessment, Brussels, Belgium.

ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals. (2010). Technical Report No. 110. Guidance on Assessment Factors to Derive a DNEL, Brussels, Belgium.

ECHA, European Chemicals Agency. (2012). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1, November 2012, https://echa.europa.eu/documents/10162/13632/information_requirements_r8_en.pdf .

ECHA, European Chemicals Agency. (2017). Guidance on the Biocidal Products Regulation. Volume III Human Health - Assessment & Evaluation.(Parts B+C). Version 4.0. December 2017, https://echa.europa.eu/documents/10162/23036412/biocides_guidance_human_health_ra_iii_part_bc_en.pdf/30d53d7d-9723-7db4-357a-ca68739f5094.

EFSA, European Food Safety Authority. (2012). Guidance on Dermal Absorption. Scientific Opinion. EFSA Panel on Plant Protection Products and their Residues (PPR). The EFSA Journal 10(4):2665, 1-30.

Gajewska, M., Worth, A., Urani, C., Briesen, H., Schramm, K.W. (2014). Application of physiologically-based toxicokinetic modelling in oral-to-dermal extrapolation of threshold doses of cosmetic ingredients. Toxicology Letters 227, 189-202.

Geraets, L., Bessems, J.G., Zeilmaker, M.J., Bos, P.M. (2014). Human risk assessment of dermal and inhalation exposures to chemicals assessed by route-to-route extrapolation: The necessity of kinetic data. Regulatory Toxicology and Pharmacology 70, 54-64.

IGHRC, Interdepartmental Group on Health Risks from Chemicals. (2006). Guidelines on route-to-route extrapolation of toxicity data when assessing health risks of chemicals, Institute of Environment and Health Bedfordshire UK.

Schröder, K., Escher, S.E., Hoffmann-Dörr, S., Kühne, R., Simetska, N., Mangelsdorf, I. (2016). Evaluation of route-to-route extrapolation factors based on assessment of repeated dose toxicity studies compiled in the database RepDose®. Toxicology Letters 261, 32-40.

SCOEL, Scientific Committee for Occupational Exposure Limits. (2017). Methodology for derivation of occupational exposure limits of chemical agents. The General Decision-Making Framework of the Scientific Committee on Occupational Exposure Limits (SCOEL) 2017, European Commission. Directorate-General for Employment, Social Affairs and Inclusion. Directorate B - Employment. Unit B.3 - Health and safety, Brussels. https://publications.europa.eu/en/publication-detail/-/publication/3c8ef3e0-48fc-11e8-be1d-01aa75ed71a1/language-en.

Stokinger, H.E., Woodward, R.L. (1958). Toxicologic methods for establishing drinking water standards. Journal of the American Water Works Association 50, 515-529.

Sweeney, L.M., Gargas, M.L. (2016). Route-to-route extrapolation of 1,2-dichloroethane studies from the oral route to inhalation using physiologically based pharmacokinetic models. Regulatory Toxicology and Pharmacology 81, 468-479.

WHO, World Health Organization. (2010). IPCS Harmonization Project Document No. 9. Characterization and Application of Physiologically Based Pharmacokinetic Models in Risk Assessment, http://www.who.int/ipcs/methods/harmonization/areas/pbpk_models.pdf.

REPORT 6:
Time Extrapolation

RESEARCH PROJECT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Summary

Several attempts were made in the past to obtain empirical information on the quantitative differences between critical doses observed in shorter-term and longer-term experimental animal studies. These empirical data evaluations should help to assess which assessment factor is suitable to estimate an equivalent dose for chronic exposure, when starting with a point of departure from a subacute or subchronic study. To update and increase this empirical database we evaluated two large datasets: NTP studies and repeated dose studies from REACH registrations.

Based on the studies performed by NTP, out of all available NTP reports up to TR-596, those who fulfilled the study selection criteria were manually evaluated. For each evaluated report, doses (or concentrations in case of inhalation studies) corresponding to the NOAEL and/or LOAEL were determined for each type of study (species, sex, study duration) and type of endpoint (bodyweight, local and systemic effects). In case of REACH data, these doses were determined based on structured data provided by the European Chemicals Agency (ECHA), which was extracted from the REACH IUCLID database. The REACH database contains considerably more repeated dose toxicity studies than the NTP reports, however strict selection criteria, based on the study metadata reported in IUCLID, were necessary to ensure sufficient quality of the data used. Using the obtained NTP and REACH datasets, for each substance the available subacute/subchronic, subchronic/chronic and subacute/chronic study pairs were compared by calculating the ratios $dose_{shorter\ study}/dose_{longer\ study}$. The resulting ratio distributions were further stratified according to study parameters to evaluate possible influencing factors. This evaluation of both datasets led to the conclusion that no consistent differences with regard to time extrapolation for the variables

- Route of application (oral, inhalation)
- Sex
- Species
- Toxicity endpoints after inhalation (local, systemic)
- Target organs
- Substance classes (exemplary examined for two groups of substances with NTP data)

are evident. Due to observations on reporting quality of REACH data and the restrictions of a largely automated study evaluation (necessary due to the initially high number of studies), we consider the REACH database less reliable than the manual evaluation of NTP data. This conclusion is supported by a larger GSD (geometric standard deviation) for REACH data compared to NTP data, pointing to higher variability in this dataset. Therefore, we conclude that the combined dataset of ratios from oral and inhalation NTP data is adequate for proposing distributions for time extrapolation. The data are derived from studies on 256 substances, from which close to 400 (subacute/subchronic and subacute/chronic each) or more than 1200 (subchronic/chronic) ratios were calculated. For local effects in the respiratory tract after inhalation, the same distributions as for systemic effects are proposed, a conclusion which is supported by publications pointing to similar or higher ratios for local compared to systemic effects.

This report presents one of the few data evaluations covering both sub-steps (subacute/subchronic and subchronic/chronic**)** as well as the full span (subacute –

chronic) of the most frequently used time extrapolation steps. The GMs (geometric means) of ratios obtained for subacute versus chronic (GM 4.11) and subacute versus subchronic (GM 1.60) NTP studies fit very well to other evaluations published in recent years. The ratios for subchronic versus chronic studies (GM 2.93) are at the upper end of the range reported in recent evaluations. However, multiplication of GMs or medians of the two sub-steps yield values in agreement with the subacute – chronic ratios, indicating consistency in the three datasets.

Abbreviations

BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
BMD(C)	Benchmark dose (concentration)
BMD(C)L	Benchmark dose (concentration) lower bound
BMDU	Benchmark dose upper bound
CAS	Chemical Abstracts Service
CI	Confidence Interval
CSV	Comma separated values
ECHA	European Chemicals Agency
EF/AF	Extrapolation factor /Assessment factor
EPA	Environmental Protection Agency (in the US)
ESR	Endpoint study record
GM	Geometric mean
GSD	Geometric standard deviation
IUCLID	International Uniform ChemicaL Information Database
KS	Klimisch reliability score
LOAEC	Lowest observed adverse effect concentration
LOAEL	Lowest observed adverse effect level
NOAEC	No observed adverse effect concentration

NOAEL	No observed adverse effect level
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
PBPK	Physiology-based pharmacokinetic (model)
POD	Point of departure
PPP	Plant protection products
QSAR	Quantitative structure activity relationship
RAC	Committee for Risk Assessment
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals, Regulation (EC) No 1907/2006
SMILES	Simplified Molecular Input Line Entry Specification
TR	Technical report
UUID	Universally Unique Identifier

1 Introduction

Occupational exposure limits are health-based values aiming to protect against adverse effects at the workplace under long-term exposure conditions. Ideally, they are derived from (either human or experimental animal) data representing chronic exposure conditions, as it is assumed that effects may aggravate and/or critical exposure levels may decrease with prolongation of exposure. However, toxicological data covering chronic exposure conditions are often absent. Therefore, several attempts were made in the past to obtain empirical information on the quantitative differences between critical doses observed in shorter-term and longer-term studies (see Kalberlah and Schneider 1998 for an overview on evaluations up to 1996). Such empirical datasets should help to assess which assessment factor is suitable to estimate an equivalent dose, when starting with a point of departure from a subacute or subchronic study. As a rule, assessment factors should only be applied where substance-specific information is lacking. Substance-specific information might consist of, e.g. data from a PBPK model estimating steady-state concentrations under chronic exposure conditions or information on exposure-time dependency of effects from related substances.

Here we present the methodology and results for an evaluation of two large databases, the set of studies published by the US National Toxicology Program (NTP) and studies submitted to the European Chemicals Agency as part of registration dossiers (sections 2 and 3). In section 4 we discuss the methodological issues and compare the results with other investigations of the last twenty years on this subject.

2 Methods

2.1 Selection and analysis of NTP study data

2.1.1 Criteria for study selection

To establish a database for derivation of assessment factors regarding

- Time (exposure duration) and
- Interspecies extrapolation

NTP reports were selected for analysis according to the following criteria:

- Only studies with inhalation or oral exposure were included
- Studies, which are only available as draft reports, are excluded
- For any substance two studies of different exposure duration have to be reported. Occasionally, for „technical reports“, which only report a 2-year study, a corresponding „toxicology report“, which contains the shorter study periods is available. Knowledge of this scenario entails evaluation of both studies, but this was not checked systematically.
- Studies, where no effects were specified according to the criteria outlined below, are excluded

The list of all NTP technical reports (TR) is available at: https://ntp.niehs.nih.gov/data/tr/index.html. On 17.1.2020, 596 technical reports were listed on this website, with report TR-596 from 2018 being the newest one (not including drafts).
Around the year 1985 (TR-255) the reports start to include 13-week studies, in addition to the 2-year carcinogenicity studies. In reports older than TR-255, 13-week-studies are less frequent. Going further back in time the investigation depth of the subchronic studies deteriorates, so that evaluation of reports older than about TR-200 (around the year 1980) is not useful. Sighting of NTP studies to check for matches with the selection criteria was stopped at TR-184, with TR-195 being the last study where effects were evaluated because it passed the criteria.

2.1.2 Procedure to specify NOAEL and LOAEL for the analysis and to establish the dataset

For each evaluated NTP report, the following study parameters were documented:

- Substance name and CAS number
- TR number and year of publication
- Exposure route and type: inhalation, oral (gavage), oral (feed), oral (water)
- Dose levels, for each study length, species and sex

- Unit of dose levels. For approximately half of the reports with exposure via food or water, a bodyweight dose is not readily available from the reports for all study durations. To allow a meaningful interspecies comparison, food factors (based on assumptions for daily food and water uptake) according to Table R.8-17 in the ECHA Guidance R.8 are used to derive a dose in mg/kg bodyweight for these studies.

The depth of subacute studies differs considerably from the longer studies in terms of investigated organ histopathology. This means that the absence of reported effects does not imply the actual absence of toxicity, which is a potential confounder for the downstream analysis. In consequence, for the 2-week studies only effects on body weight were evaluated (see also **Table 2-1**).

Table 2-1 Evaluated endpoints by study type

Study type	**Body weight**	**Local effects in the respiratory tract (only for inhalation studies)**	**Systemic effects**
2 weeks	X		
13 weeks	X	X	X
2 years	X	X	X

For each evaluated endpoint, a NOAEL and a LOAEL were identified from the study summaries and consultation of respective sections of the main report. Generally, it was assumed, that relevant adverse effects are reported in the main report. The appendix was only consulted in exceptional cases.

For specification of NOAEL and LOAEL and for building the dataset for analysis, the following rules were applied:

For all types of effects:

- Doses leading to severe mortality (>= 80%) in the subacute and subchronic studies were flagged (by putting this dose in parentheses) and not further considered in the analysis. All doses below this threshold were used in the analysis. For chronic studies, mortality was no evaluation criterion.
- If a NOAEL or LOAEL could not be properly defined (because of the LOAEL being the lowest tested concentration or the NOAEL being the highest evaluated concentration), the corresponding dose was documented as “n.a.” (“not available”) to allow easy handling of this case in the downstream analysis.
- If a study type was not performed or reported by NTP, it is documented as “n.t.” (“not tested”).
- If a study type was repeated and both studies are reported in the report, generally only the second study was evaluated.

- When statistics were given, effects were only considered if they were statistically significant.

For effects on body weight:

- A change of 10% at the end of the study period was considered an effect. In borderline cases, the statistical evaluation or interpretation of the study directors was followed, if available.
- Effects on body weight were not documented in case of indications that a body weight reduction is a mere consequence of poor palatability of the substance in feed or water.

For local and systemic effects:

- The target organs for effects determining the LOAEL were documented with the goal to be used for organ specific data analysis. Due to the limited data, this is meaningful only for the most common organs (e.g. liver, kidney). However, also rare target organs were documented, where possible.
- Local effects were only evaluated for inhalation studies, i.e. lesions in the gastro-intestinal tract were not evaluated for oral studies. In case of feed studies, nasal lesions were considered to be local effects from direct contact with the substance, except when the lesions can be ascribed to a systemic mechanism.
- Effects on organ weights were not documented, as they are generally not reported for 2-year studies (where the focus is on the histopathological evaluation) and inclusion of organ weight effects in 90-day studies would introduce a bias towards lower effect concentrations in the subchronic studies.
- Reproductive parameters, where the same reasoning applies (primarily sperm motility and oestrous cyclicity), were also excluded from documentation.
- Only non-neoplastic lesions were considered, neoplasms were not documented as effects. Some of the reported lesions are part of a morphological continuum between a non-neoplastic and neoplastic phenotype. As a general rule, hyperplasia and metaplasia were documented by us.
- In older NTP reports, it has been observed that positioning of cages (i.e. proximity to the room lighting) might have been responsible for observed eye lesions. The first report where positioning was explicitly taken into consideration by NTP when evaluating effects on eyes was TR-368 (1989). If a study was older than this report, lesions to the eye were not documented by us.
- Clinical parameters (e.g. T3 or T4 hormone levels) were not considered.
- Changes in the severity of an effect (often classified on a scale from 0 to 3 or 4) were considered as critical effects if the change in severity is either described in the text or a dose response is unequivocally evident from the result tables, even when a statistical evaluation was not performed by NTP.
- After inhalation, lesions in the bronchial lymph nodes were considered local effects, but lesions in the mediastinal lymph node were considered systemic.
- Effects which are clearly not considered adverse by NTP were not documented.
- Potentially controversial effects which we generally considered adverse:

  - tissue pigmentation
  - tissue regeneration
- Potentially controversial effects which we generally did not consider as critical effects:
  - Respiratory effects in conjunction with Sendai virus infections
  - Effects influenced by *Klebsiella pneumoniae* infections

2.1.3 Derivation of factors for time and species extrapolation

This section explains how ratios and their distributions regarding time extrapolation (the topic of this report) were derived. The procedure to derive the distributions for species extrapolation (the topic of a separate report) is highly overlapping and is also described here.

For the whole documented dataset, for any pairs of

- 2 study types of different lengths (but with same species)
- 2 study types of different species (but with same length)

the ratios $dose_{study1}/dose_{study2}$ were derived. For comparison of two studies of different duration, study1 and study2 are defined as study1 having a shorter study duration than study 2. In consequence, if a substance has a lower effective dose at longer exposure times, the ratio for this study comparison will be greater than 1. For species comparisons, the doses were divided in the way that the species of study1 has a larger bodyweight than species from study 2.
The ideal case for a dose comparison is when both compared studies have a numerically defined NOAEL and LOAEL. If a NOAEL is not defined for an endpoint, it follows that the corresponding LOAEL could have been lower, in case a lower concentration had been tested. Thus, this LOAEL carries a higher uncertainty than a LOAEL where the corresponding NOAEL is defined (as in the latter case any dose between the NOAEL and LOAEL could also be a LOAEL, but a dose below the NOAEL cannot). A corresponding scenario exists when the LOAEL is not defined for an endpoint. Indeed, for a considerable number of evaluated study endpoints, no NOAEL or no LOAEL could be derived within the tested dose range. If a ratio is to be computed from one or even two such study endpoints, several cases with different qualities of uncertainty regarding the computed ratio occur. From some of these cases, meaningful information on the ratio of the effective doses can be drawn despite the inherently higher uncertainty. How these cases were handled is explained here case-by-case in more detail and summarised in **Table 2-2**.
In the **ideal case (#1 in table)**, the ratio is derived based on the NOAELs of both studies. Depending on dose spacing this might lead to different ratios than a ratio derived from LOAELs. To evaluate the impact of this decision, for this dataset the distribution of NOAEL-ratios was compared to the distribution of LOAEL-ratios. No difference in the distributions was apparent.

If no value is available for **1 of the 4** dose descriptors $NOAEL_{study1}$, $NOAEL_{study2}$, $LOAEL_{study1}$, $LOAEL_{study2}$, two subcases have to be considered:

1. If no value is known for the $NOAEL_{study1}$ or $LOAEL_{study2}$, a ratio can be calculated based on the type of dose descriptor which is available for study1 and study2. This ratio represents a **maximum** for the chosen dose spacing **(#2 in table)**. In other words, if more doses had been tested to achieve the "ideal case", the calculated ratio could only be less or equal.
2. The other way round: if no value is available for the $NOAEL_{study2}$ or $LOAEL_{study1}$, a ratio is calculated which represents a **minimum (#3 in table)**.

If **2 of the 4 dose** descriptors have no value, three subcases occur:

1. If the NOAEL of both studies or the LOAEL of both studies is unknown, no meaningful ratio can be calculated as the numerator and denominator are unbounded **(#6 in table)**.
2. If $NOAEL_{study1}$ & $LOAEL_{study2}$ are unknown, a ratio can be derived which represents a maximum, with the same explanation as for the 1 of 4 case, but with a higher likelihood that the ratio would be lower with additional doses **(#4 in table)**.
3. Conversely, the same applies the other way round when $NOAEL_{study2}$ & $LOAEL_{study1}$ are unknown. This represents a minimum ratio **(#5 in table)**.

During development of the methodology, it was tested whether inclusion of these maximum or minimum ratios has an influence on the ratio distribution. Inclusion of these ratios generally produced broader distributions. However importantly, the ratios which shifted the distributions towards lower values predominantly were maximum values, and conversely the minimum values were overrepresented in the upper quantiles of the distributions. This means that inclusion of the minimum and maximum ratios better reflects the distributions which would have been derived had there been a larger dose range. It was concluded to **include such ratios in the distributions if they belong to the cases with lower uncertainty (case #2 and #3)**. The cases where a minimum or maximum can be derived, but with a higher uncertainty **(case #4 and #5) are not used** for further analyses. Depending on the compared conditions, the fraction of discarded ratios based on case #4 and #5 are between 42.8 and 48.3%[1] of all calculated ratios. For **case #6 no ratios** were calculated.

[1] Fractions of discarded ratios with the highest uncertainty (case #4 and #5): subacute/subchronic: 44.9 %; subacute/chronic: 44.7%; subchronic/chronic 42.8%; interspecies ratios (rat/mouse): 48.3%

Table 2-2 Procedure to derive ratios from NTP studies, depending on NOAEL and LOAEL availability. See text for more detailed explanation.

Case #	Description	Undefined NOAEL/LOAEL	Ratio derived as	Remarks
1	Ideal case	None	$NOAEL_{study1}$/ $NOAEL_{study2}$	A ratio could also be derived based on LOAELs, no apparent difference between the two distributions was observed
2	1/4 undefined	$NOAEL_{study1}$	$LOAEL_{study1}$/ $LOAEL_{study2}$	Ratio represents a maximum, because $LOAEL_{study1}$ could only be smaller if lower doses would have been tested and $NOAEL_{study2}$ could only be higher if higher doses would have been tested
		$LOAEL_{study2}$	$NOAEL_{study1}$/ $NOAEL_{study2}$	
3	1/4 undefined	$NOAEL_{study2}$	$LOAEL_{study1}$/ $LOAEL_{study2}$	Ratio represents a minimum, because $LOAEL_{study2}$ could only be smaller, if lower doses would have been tested and $NOAEL_{study1}$ could only be higher if higher doses would have been tested
		$LOAEL_{study1}$	$NOAEL_{study1}$/ $NOAEL_{study2}$	
4	2/4 defined	$NOAEL_{study1}$ & $LOAEL_{study2}$	Not used	Ratio represents a maximum; Not used because of high uncertainty
5	2/4 defined	$NOAEL_{study2}$ & $LOAEL_{study1}$	Not used	Ratio represents a minimum; Not used because of high uncertainty
6	2/4 defined	$NOAEL_{study1}$ & $NOAEL_{study2}$ or $LOAEL_{study1}$ & $LOAEL_{study2}$	Not calculated	Numerator and denominator are unbounded. Not used because no meaningful ratio can be calculated

The ratios were derived for each sex separately and it was investigated whether the relevant characteristics of the distributions changes by pooling the data from male and female animals. Other experimental factors like the compared endpoints (bodyweight, local effects and systemic effects), species (for time extrapolation) or compared duration (for interspecies extrapolation) were also evaluated regarding their influence on the resulting values.

2.1.4 Data visualization and statistical evaluation

The distributions were usually visualized as a composite plot consisting of three graphical representations of the data. From top to bottom the representations are the density of the distribution, a dot plot with one dot (jittered vertically for visibility) for each ratio and a boxplot (**Figure 2-1**). The plots were generated with the standard settings of the R package 'ggplot2'[2], with the following exceptions. The kernel bandwidth was

[2] More precisely, the plot element representing the density was generated by the geom 'geom_half_violin', which at the time of writing was not included in 'ggplot2', but provided by the package 'gghalves'.

adjusted by the factor 2 in order to smooth out peaks which arise from overrepresented ratios because of intentionally chosen doses (which are usually rounded). Also, the trimming of the density at the position of the minimum and maximum value was disabled. For the boxplot, outliers were omitted but for all remaining settings the standard values were used, i.e. the hinges represent the first and third quartile and the whiskers extend to the highest value which is in the 1.5-fold inter-quartile range from these quartiles. All values outside this range are not plotted (but they are included in the dot plot).

Values are $\log_{10}$-transformed for plotting. As the encountered distributions follow a lognormal distribution reasonably well, this transformation provides a better impression of the characteristics of the distributions. Values in the text and tables are on the normal scale, except explicitly stated otherwise.
Statistical evaluation of differences between distributions was performed using the bootstrap method on chosen parameters of the distributions. Comparing certain parameters, instead of the whole distribution, is intentional as the later usage of the derived distributions will primarily involve only few parameters and most of the features are of lower relevance. It is not possible to determine which parameters will ultimately be most relevant, but we think comparing the geometric mean and the 75%-percentile provides a reasonable coverage of possible use cases. CI of the chosen parameters were analysed by using the function ‘boot.ci’ from the R package ‘boot’ with 10000-fold resampling and type = ‘perc’. A parameter was considered different between compared distributions if the 95% CI don’t overlap.

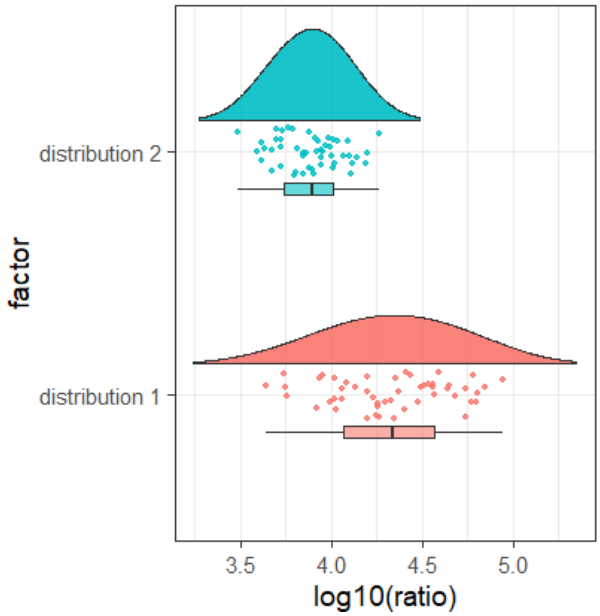


Figure 2-1 Example of the visualization used for the derived distributions. From top to bottom, each compared distribution is represented by a density estimation, a dot plot and a box plot. Values are log-transformed for plotting

2.2 Processing of REACH study data

2.2.1 General approach

Study data on repeated dose toxicity for the oral and inhalation route was provided by ECHA, i.e. information from the endpoint study records (ESR) of section 7.5.1 and 7.5.2 of all registration dossiers, upon a request from FoBiG and BAuA. In this request FoBiG identified those fields containing relevant information for this evaluation (see below). The data was exported from the confidential database on 23.10.2018. According to ECHA, no distinction between confidential and non-confidential data can be made for the data export.

The data was provided by ECHA in the form of Excel® files. One file (from here on referred to as "**general study data**") contained information on the subsections 'Administrative data', 'data source', 'materials and methods – test guideline' and 'materials and methods – test material' for each ESR of both routes. Further data was received in separate Excel® files for each route, as the variable names differ between oral and inhalation exposure: two files containing information on the subsection 'Test animals' and 'Administration / exposure' of each ESR, from here on referred to as "**exposure data**" and two files containing information on the subsection 'effect levels' of each ESR, from here on referred to as "**effect level data**". In addition, a file containing the data from IUCLID subsection 'Target system / organ toxicity' for both routes (referred to as "**target organ data**") was provided by ECHA as a CSV (comma-separated values) file at a later stage of the project (exported on 4.7.2019).

All these files contain the dossier UUID (universally unique identifier) and ESR UUID, the combination of both being unique for each study record of a registered substance. The combination of these UUIDs is used for joining data in the data cleaning process. For clarification, the ESR UUID alone is not sufficient to identify an ESR, as it might be identical across several dossiers in case of a read-across. A flowchart of the data processing is shown in **Figure 2-2**:

The four data sets ("general study data", "exposure data", "effect level data", target organ data") are curated and filtered to obtain the relevant variables in appropriate data types for the analysis. The dossier UUID and study record UUID is used to identify unique studies in the datasets. Toxicologically relevant relations between the variables are used during data curation, exemplary shown in **Figure 2-2** on two examples: three different variables are used to derive the study duration (described in detail in section 2.2.3.5) and a single numerical value for the dose descriptor is derived from the information in the effect level table (details in 2.2.3.10).

All files received as Excel® files were converted to the CSV format using Excel® and afterwards R was used for all further handling of the data. All processing was performed using documented code in R with the aim to achieve reproducibility. The description of the methodology in this section focuses on conveying the criteria and principles used during data cleaning. For reasons of confidentiality it is not possible to provide the raw data.

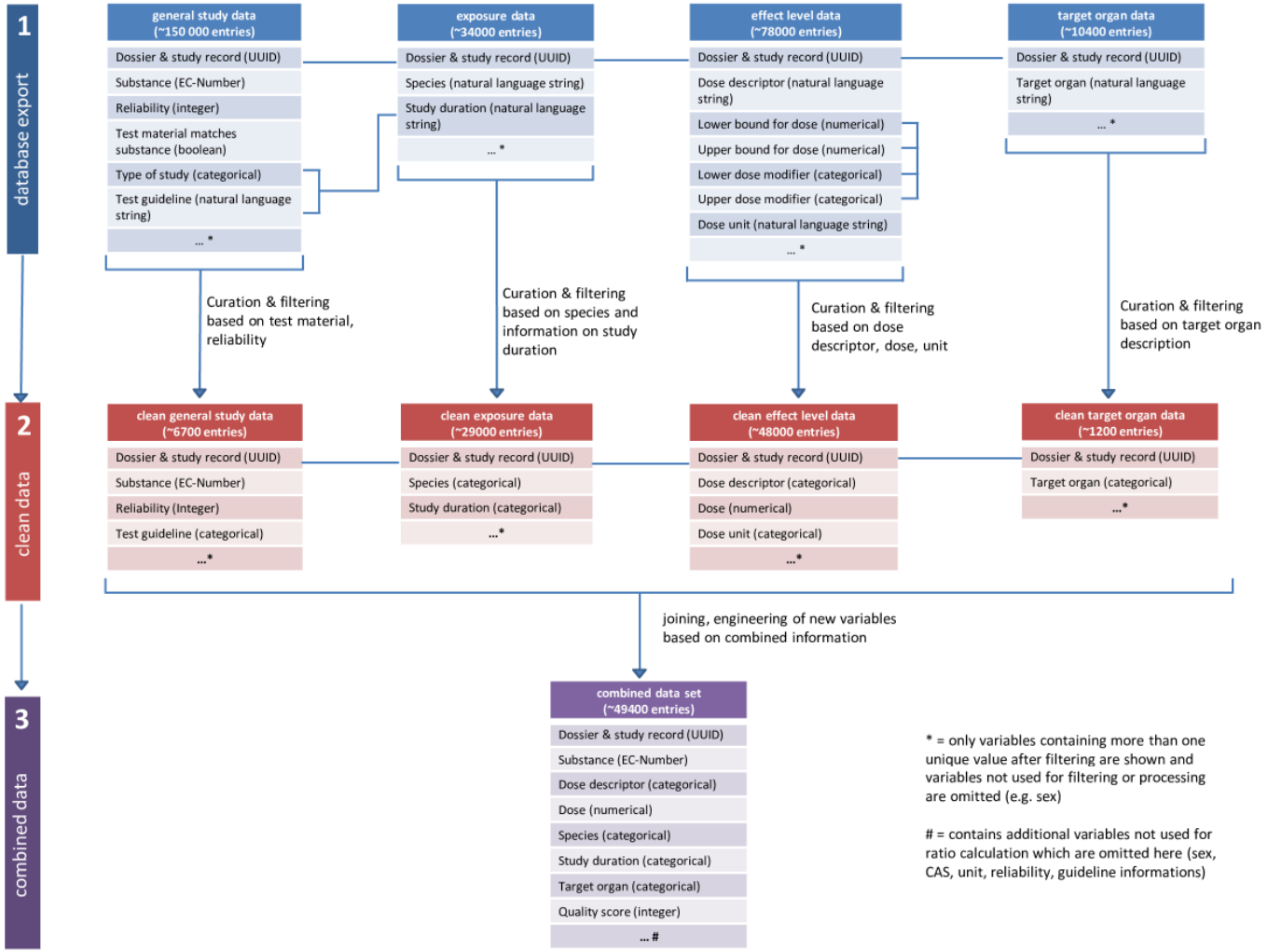


Figure 2-2 Flowchart of the data cleaning process

2.2.2 Selection criteria

The quality of data coming from the ECHA database is varying significantly, depending on the study itself but also on the way the registrants entered the data in IUCLID. The following selection criteria were defined to provide a sufficiently reliable database for the later analysis of dose ratios:

- The study has a reported reliability of 1 or 2 (Klimisch score)
- The test material in the study matches the registered substance
- Only experimental studies, no read-across or QSARs
- Studies for which an inappropriate guideline was reported are excluded
- Only studies where the study duration can be categorized as either “subacute”, “subchronic” or “chronic”
- Only studies where the tested species can unambiguously be determined
- Only dose values given for dose descriptors (e.g. NOAEL, LOAEL) suitable for comparison or a descriptor that can reliably be converted to such (see 2.2.3.7)
- Only dose values given in a unit suitable for comparison or a unit that can reliably be converted to such (see 2.2.3.8)
- For dose values which are given as a range or unbounded, a single value is chosen (see 2.2.3.10)

2.2.3 Data cleaning

For several selection criteria extensive data cleaning was necessary. Although much of the relevant data is stored as categorical variables derived from the picklists in IUCLID, registrants can often make use of the “other:” category which is then further specified in “natural language”. Natural language refers to the language how it is used and understood by humans, which in many cases is not readily understandable by computers. Sometimes, the categories of a variable are not sufficiently specific to serve the purpose of this analysis, e.g. the duration of a study may be determined by the variable ‘endpoint’ in the section ‘administrative data’ which takes categorical values corresponding to “subacute”, “subchronic”, “chronic” and “repeated dose”, while “repeated dose” could mean any of the former options. A pragmatic solution would be to simply discard all data points with ambiguous values, but this was deemed to result in too much loss of information. Therefore, ambiguous values were converted into useful categorical values where possible. Another issue addressed during data cleaning is resolving conflicting data from supposedly wrong entries, e.g. a study filed as a chronic study but performed according to a subacute guideline. In the following, more details on the cleaning steps are provided.
Applying the cleaning steps as described in the following and removing all studies from the dataset which did not fulfil the quality criteria reduced the initial set of 150000 study records (with potentially multiple reported dose descriptors) to about 49400 dose descriptors from roughly 6700 study records which form the basis for ratio comparisons.

2.2.3.1 Reliability

Reliability is directly available in the data as numerical values. The submitted reliability by the registrants was used without running plausibility checks. Implementing plausibility checks is possible but time consuming. More importantly, plausibility checks on reliability are unlikely to be more effective in removing low quality studies than the following selection criteria.

2.2.3.2 Check whether test material matches the registered substance

Only studies which were actually performed with the registered substance were retained for further processing. This information is not explicitly provided by the registrants; instead, it was derived by ECHA based on structured information on the registered and the tested substance. For some 500 study records, structured information on the test material is not available. To handle these cases, ECHA engineered an additional feature, where the required information is derived by other means. The exact implementation of the two classifiers by ECHA is unknown to us.

2.2.3.3 Only experimental studies

The information in the field ‘Type of information’ from ‘administrative data’ needs to be ‘experimental data’. This step removes unsuitable study records like information derived from QSAR, calculations, planned studies and read-across data, which were not caught by the preceding filter step.

2.2.3.4 Testing guideline

All studies without information regarding guideline whatsoever were removed. This is a very conservative filtering step, which provides a higher confidence in processing of the remaining studies, but potentially lead to the loss of experimental data that would be valid for the downstream analysis. For example, among the roughly 3000 study records removed by this step, about 100 are reported as key studies with a reliability of 1. These entries were checked in detail to determine whether these cases could be handled more appropriately. For about half of the studies in question the reliability seemed unjustified. The other half seemed like studies of good quality where the registrants did not provide a description of the guideline, but also contained studies which would have been filtered out by later steps (e.g. 5-day short term inhalation studies). Taken together, it was deemed too time-consuming to correctly identify the studies which should be kept in the dataset at this point.
After this filtering step, the testing guideline was extracted from the given information and studies following a guideline deemed unsuitable for the analysis (according to the negative list in **Table 2-3**) are discarded. Of note, this list was constructed to exhaustively handle testing guidelines occurring in the data sets at this point of processing; it is not exhaustive for all guidelines registrants may possibly submit.

Table 2-3 Negative list for testing guidelines

Guideline		**Reason for exclusion**
OECD	401, 403, 425	Acute study
	408 (for inhalation studies only)	Oral study
	478	Genotoxicity
	414, 415, 416, 421	Reproductive & Developmental
	419, 424	Neurotoxicity
	440	Uterotrophic assay
EU Method	B.38, B.43	Neurotoxicity
EPA OCSPP	870.3550	Reproductive & Developmental
	870.6xxx series	Neurotoxicity
EPA OPP	81-7, 82-5, 82-6	Neurotoxicity
EPA PPT	798.6050	Neurotoxicity

2.2.3.5 Study duration

Information on the duration of a study can be found in three different variables in the IUCLID data format:

1. The ‘endpoint’ variable under ‘administrative data’
2. The test guideline
3. The ‘duration of treatment’ variable under ‘Administration / exposure’, which for inhalation studies is, understandably, often confused with the variable ‘frequency of treatment’

A tiered approach was used here. On the data with oral exposure (after preceding cleaning steps, this corresponds to 5204 records out of 6710 for both exposure routes) it was empirically checked whether derivation of the study duration from the administrative data or from the guideline leads to fewer misclassifications. For about 80% of study records, both approaches may be used on their own to derive an unambiguous, but not necessarily correct, category for the study duration. However, for about 6% of those 80% (corresponding to 281 records), the derived category was divergent for the two approaches (e.g. the study is reported as a chronic study in the administrative data, but a guideline for a subchronic test is given). For these cases, the duration derived from the testing guideline was better matching the information under ‘Administration / exposure’ than the duration derived from the administrative data.

Accordingly, the study duration was first derived from the testing guideline (see 2.2.3.4) as given in **Table 2-4**. When this was not possible, the study duration was derived from the administrative data.

For those cases, where the two steps above failed, the study duration was extracted from the 'duration of treatment' field, handling different abbreviations, languages and units in an iterative process using the criteria as given in **Table 2-5**[3]. This method for classification is the most error-prone of this tiered approach, because in cases like "exposure for 14 days (male), at least 28 days (females)" or "6 h/day, 5 days/week, 13 weeks" the correct extraction of study duration is not trivial. Therefore, all classifications which needed to be performed according to the last step in the procedure (applicable for 67 oral studies, 8 inhalation studies) were manually checked for correctness based on the available information. After this last step only 1 inhalation and 3 oral studies could not be categorized. For these, the duration category (subacute, subchronic, chronic) was manually determined by evaluating all available information in free text fields of these studies, or in case that was not possible, the study was excluded from the analysis.

[3] Please note that no paramount definition of class boundaries for study durations exist and that the boundaries in Table 2-5 are to a certain extent arbitrary. As explained in the discussion, the exact location of the boundaries has negligible impact on the analysis.

Table 2-4 Categorization for study duration by the given testing guideline. Listed are only guidelines which actually occurred in the dataset after the preceding filter steps

Guideline		**Time category**
OECD	422	subacute
	407, 412	subacute (specifically 28 days)
	408, 413	subchronic (specifically 90 days)
	451, 452, 453	chronic
EU Method	B.7, B.8	subacute (specifically 28 days)
	B.26, B.27, B.29	subchronic (specifically 90 days)
	B.30, B.32, B.33	chronic
EPA OCSPP	870.3050, 870.3650	subacute
	870.3150, 870.3465, 870.3700	subchronic
	870.4100, 870.4200, 870.4300	chronic
EPA OPPT	795.2600, 798.2450, 798.2650	subchronic (specifically 90 days)
	798.3260, 798.3300	chronic
EPA OPP	82-1, 82-4	subchronic
	83-1, 83-5	chronic

Table 2-5 Category boundaries for study durations[3]

Study duration in days	**Time category**
7 - 70	subacute
71 – 180	subchronic
> 180	chronic

2.2.3.6 Species

The species which was used in the experimental study can be defined in IUCLID either by picking pre-defined categorical values or by picking the “other:” option with further specification in natural language. Study records with species defined as “other:” were converted to picklist values where possible. Occasionally, multiple species are given in the specification. These records were discarded as it was deemed unfeasible to properly decipher which of the linked dose values belong to which species. This cleaning step was performed with the extracted file on exposure data.

2.2.3.7 Dose descriptors

This was the first step of cleaning the “effect level” data. Each study record may be linked to multiple data points with effect levels. First, all data points without any information on the dose descriptor were removed (51100/16900 left). Next, all dose values reported as BMD05, BMDL05, BMC05 or BMCL05 (together only 7 data points) are discarded. Dose values given as BMDL10 or BMD:[4] (together 147 oral data points) or BMCL10 (59 inhalation data points) were kept in the dataset (and treated equivalent to NOAELs respectively NOAECs). About 1300 (oral) and 1270 (inhalation) dose descriptors are reported as “other:” with a clarification given in natural language. Much effort has been put into classifying these entries as NOAEL or LOAEL by manually going through a list of all occurring entries. For example, the description “no toxic effect level” was considered a NOAEL or “Minimal-Observed-Adverse-Effect Concentration (MOAEC)” was considered a LOAEC. For the sake of easier data processing and visualization, it was not differentiated between “concentration” and “level” descriptors, e.g. “NOAEC” was replaced by “NOAEL”. NOELs and LOELs likely have too little data points to be of value but were kept in the dataset. Further in this report, the numerical values reported as NOAEL, LOAEL, NOEL, LOEL, BMD etc. were termed as “dose values”.

2.2.3.8 Cleaning units

Doses and concentrations may be given in several units. The most important units are coded as categories in IUCLID but may also be reported in non-standard units as “other:” followed by a clarification in natural language. For oral studies, only units which refer to a daily bodyweight dose were retained in the dataset. The number of studies which reported exclusively a concentration in food or water was low and consequently dose descriptors with units describing a concentration in feed or water were discarded. All remaining records had the unit mg/kg bw.

For inhalation studies, the common usage of “ppm” (molar concentration) and “mg/m³” (weight concentration) made additional conversion steps necessary. First, units of weight concentrations other than mg/m^3 (e.g. µg/m^3) were converted to mg/m^3 where

[4] The IUCLID file definition provides BMD05, BMDL05, BMDL10, BMC05, BMCL05, BMCL10, BMD:, BMC: as possible selections for a benchmark dose descriptor. ‘BMD:’ was interpreted as BMD10 by us, ‘BMC:’ was not encountered in the dataset.

possible and accordingly units of molar concentrations (e.g. ppb) to ppm. Then units given as “other:” were categorized to the maximal feasible extent as mg/m³ or ppm. Finally, it was tried to convert all values given in ppm to mg/m³:

- Effect level data points, where a value in both ppm and mg/m³ was available and the “basis of effect” was identical, were considered redundant and the dose value given in ppm was discarded.
- For all other cases, the molecular weight was derived from structural information (SMILES) of the tested substance. Values in ppm were converted to mg/m³ by using the rcdk package v. 3.4.7.1 (Guha 2007) and the relationship 1 [mg/m³] = 1 [ppm] * mol.weight [g/mol] * (12.19 / 293.15) [mol*mg/(g*m³*ppm)].

The redundant data points from the step before (value in ppm and mg/m³) were used for validation of this conversion. Dose values given in ppm, which could not be converted due to lack of structural information (162 values from 34 substances) were discarded.

2.2.3.9 Identifying target organs

In order to allow the investigation of organ specific distributions, information on the target organ was derived from the variable “system” and “organ” under “target system / organ toxicity” from the results and discussion section of the endpoint study records. Although registrants have various possibilities to indicate affected organs, these fields were the only identified sources of structured information on the target organ. The organ may be specified as “other:” in combination with an explanatory note in natural language. In these cases, it was attempted to categorize the study results based on the given text. In cases where several studies were available for the same study duration and species, it was assumed that a reported target organ in a single study is indicative for the same organ being a target in the remaining studies in this group. The sex of the test animals was ignored for determining the target organ, which potentially could introduce an error if a substance has a sex specific organ toxicity but the affected sex was only tested in one of the compared studies.

2.2.3.10 Ranges or unbounded values

Doses or concentrations in the REACH database are often reported as unbounded/unprecise values using quantifiers (e.g. “<”, “<=”, “ca.”) or as ranges using an upper and lower quantifier. For the purpose of the later analysis (deriving ratios of dose descriptor values), it is necessary to determine a single value without a quantifier for each effect record in the dataset. The following rules were followed to achieve this, which are best understood by looking at the examples in **Table 2-6**:

For values which are not part of a range:

1. Values specified with the quantifiers “=” or “ca.” are treated equivalent to values without a quantifier

2. All NO(A)EL with “>” or “>=” quantifiers are treated equivalent to a NO(A)EL without quantifier. Accordingly, all LO(A)EL with “<” or “<=”.
3. All NO(A)EL with “<” or “<=” quantifiers are considered not useful for the analysis and discarded. Accordingly, all LO(A)EL with “>” or “>=”.

For values which are part of a range:

4. Values specified with “ca.” in ranges are treated as “>” if the “ca.” refers to a lower bound and “<” for an upper bound.
5. Ranges starting at 0: For a LO(A)EL, this is equivalent to rule 2, and the upper bound of the range is considered the LO(A)EL. For a NO(A)EL, this is equivalent to rule 3 and the value is discarded.
6. All other ranges are converted to single values by retaining only the lower value.

Table 2-6 Examples for handling dose ranges and unbounded values

Example for rule no.	Descriptor	Value (with quantifier) before	Value after
1	LOAEL	ca. 100	100
1	LOAEL	= 100	100
2	LOAEL	< 10	10
2	NOAEL	> 10	10
3	LOAEL	> 10	-
3	NOAEL	< 10	-
2, 3	NOAEL	> 10 – < 100	10
4, 2, 3	LOAEL	ca. 10 – ca.100	100
4, 2, 3	NOAEL	ca. 10 – ca.100	10
5, (2)	LOAEL	0 – 10	10
5, (3)	NOAEL	0 – 10	-
6	LOAEL	10 – 100	10
6	NOAEL	10 – 100	10

2.2.4 Data processing

The result of the data filtering and cleaning procedure is a dataset for each route which contains in each row

- A substance, identified by CAS, EC number and name
- a single numerical dose value
- a corresponding unit, either “mg/kg bw/day” (oral) or “mg/m^3” (inhalation)
- a dose descriptor, either “NOAEL”, “LOAEL”, ”NOEL” or “LOEL”.

- the species, either “rat”, “mouse”, “dog”, “primate”, “guinea pig”, “pig”, “rabbit”, “hamster”, “human”, “miniature swine”, “cat”
- the sex, either “male”, “female”, or “male/female”
- the study duration, either “short [subacute]”, “sub” [subchronic] or “chronic”
- the reliability, either “1” or “2”
- the guideline and qualifier

In order to build ratios to investigate time and species extrapolation for any substance a single dose value needs to be selected for any unique combination of a dose descriptor, species and study duration.
The following algorithm was used to achieve this:

- **for each unique combination of substance, dose descriptor, species and study duration**
  - for each study record determine a “**quality score**”:
    - quality = 1 if Klimisch reliability score is “1”
    - quality = 2 if reliability is “2” and the guideline qualifier is “according”
    - quality = 3 if reliability is “2” and the guideline qualifier is “equivalent”
    - quality = 4 if reliability is “2” if no guideline qualifier is available
    - quality = 5 in any other case (this does not occur with the specified selection criteria above)
  - determine the **best available quality** score
  - discard all values which don’t have the best available quality
  - take the **mean** of all remaining values as final value for the unique combination

Ratios were then calculated for the time and species comparisons, differentiating between species (for time extrapolations) and study durations (for interspecies extrapolations) and dose descriptors. Regarding the dose descriptor, no differentiation was made between NOAEL and NOEL and between LOAEL and LOEL. The sex of the species was ignored when building the ratios. For those comparisons, where a ratio could be calculated based on the NOAEL as well as the LOAEL of a tested substance only one of the two ratios (i.e. the NOAEL ratios) was used in order to avoid giving data-rich substances more weight in the distributions. For the few cases where only a LOAEL-derived ratio was available, this ratio was used and in the subsequent analysis steps it was not further differentiated between NOAEL- and LOAEL-derived ratios.

3 Results

3.1 Evaluation NTP data

Ratios were calculated as explained above (section 2.1.3). In the following, the potential influence of the parameters exposure route, species and sex was analysed for each set of ratios (subacute/chronic, subacute/subchronic, subchronic/chronic) separately.

3.1.1 Subacute/chronic

3.1.1.1 Influence of exposure route

The influence of exposure route on the ratio distributions of comparisons subacute/chronic was analysed (Figure 3-1).

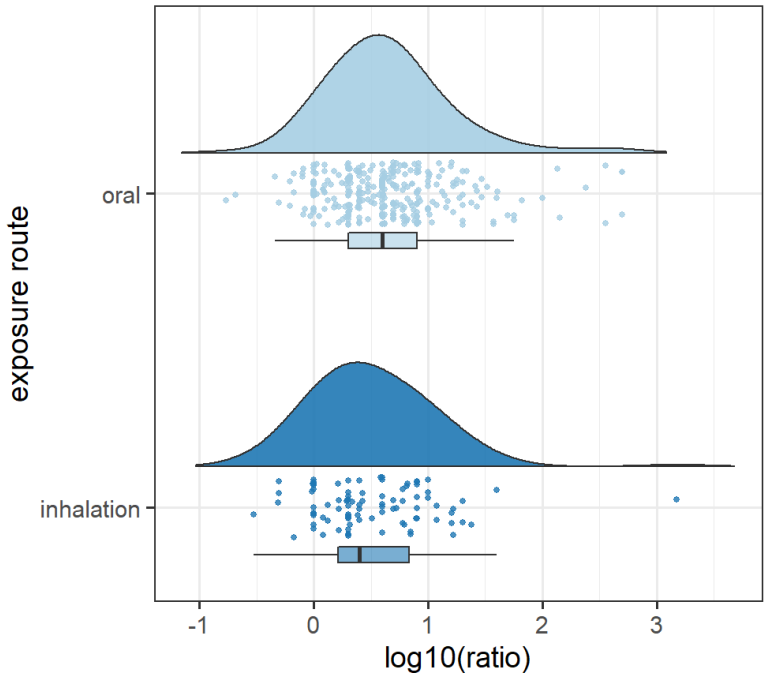


Figure 3-1: Distribution of ratios for the comparison of subacute with chronic exposure, separated by exposure route.

Exposure route might have an influence on the ratio distribution. The GM is lower for inhalation exposures than for oral, however this difference is not statistically significant (95% CI oral: 3.85 – 5.06, inhalation: 2.58 – 4.17). The dispersion is comparable between the two exposure routes (**Table 3-1**). Although the 95% CI for the 75% percentile is also overlapping (95% CI of the 75% percentile oral: 6.27 – 8.33,

inhalation: 4.67 – 8.00), route was considered relevant and data were kept separate when analysing further stratifications of the data (according to endpoints, target organs and chemical features, see sections 3.1.4 and following).

Table 3-1 Summary statistics of distributions of ratios for the comparison of subacute with chronic, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	15.75	59.52	4.40	3.41	1.00	4.00	8.00	37.50	305
inhalation	21.54	156.83	3.25	3.31	0.82	2.50	6.83	18.33	91

3.1.1.2 Influence of species

Species was also investigated as a potentially influencing factor on the distributions (**Figure 3-2**). The species-specific distributions are quite comparable in location and shape for rat and mouse (**Table 3-2**). The 95% CI for the GM have a high overlap (rat: 3.34 – 4.65, mouse 3.70 – 5.18) and also the 75% percentile is not different between the species (95% CI rats: 6.01 – 8.33, mice: 6.25 – 8.33). Because of the similar GM and spread, species was not further treated as an influencing experimental factor.

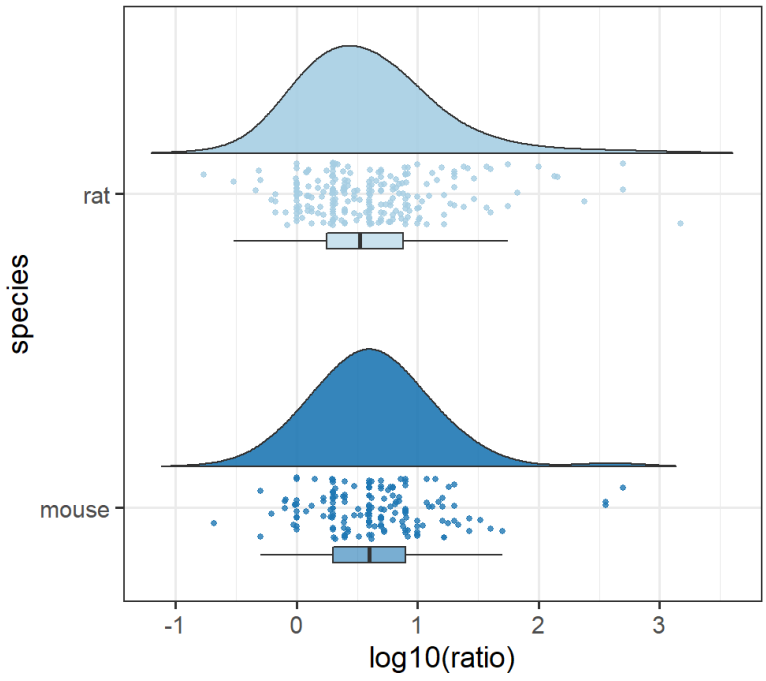


Figure 3-2 Distribution of ratios for the comparison of subacute with chronic exposure, separated by species.

Table 3-2 Summary statistics of distributions of ratios for the comparison of subacute with chronic, separated by species

Species	Mean	SD	GM	GSD	5%	Median	75%	95%	n
rat	19.62	110.25	3.93	3.64	1.00	3.33	7.55	38.75	231
mouse	13.53	54.85	4.36	3.07	0.97	4.00	8.00	21.49	165

3.1.1.3 Influence of sex

The sex-specific distributions appear to be quite comparable for male and female animals (**Figure 3-3** and **Table 3-3**). Accordingly, the 95% CI indicates that the location of the GM is not different between the two sexes at this confidence level (GM male: 2.99 – 4.16, GM female: 4.10 – 5.81). Not surprisingly, also for higher percentiles no difference between distributions was observed (e.g. the 95% CI of the 75% percentile for males: 5.71 – 8.00, for females: 6.46 – 10.14). Because the CI are overlapping at the 95% confidence level and the differences are rather small, sex was not considered as an influencing experimental factor in the subsequent stratifications.

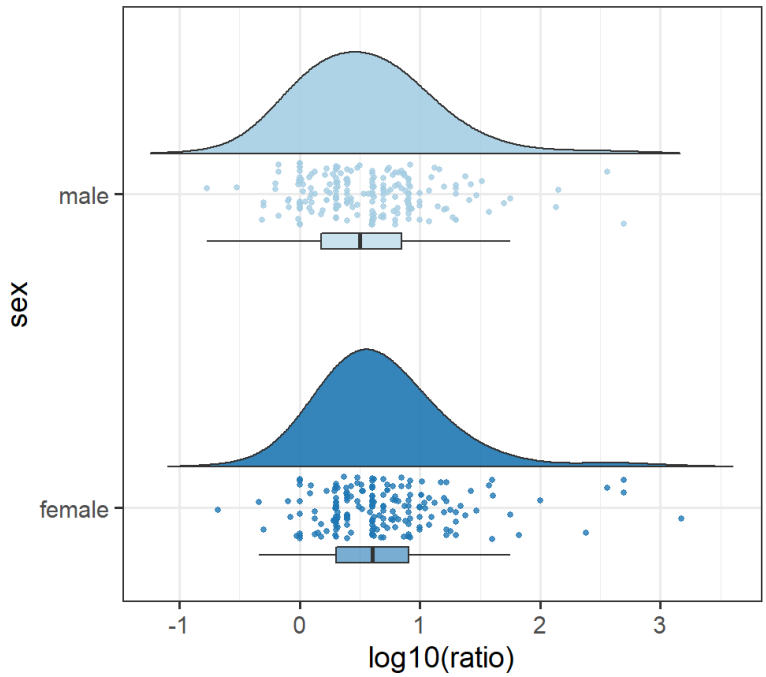


Figure 3-3 Distribution of ratios for the comparison of subacute with chronic exposure, separated by sex.

Table 3-3 Summary statistics of distributions of ratios for the comparison of subacute with chronic, separated by sex

Sex	Mean	SD	GM	GSD	5%	Median	75%	95%	n
male	11.13	44.90	3.51	3.30	0.80	3.17	7.00	25.56	205
female	23.47	122.83	4.85	3.45	1.00	4.00	8.00	40.00	191

3.1.2 Subacute/subchronic

3.1.2.1 Influence of exposure route

All parameters for the ratio distributions are quite comparable between oral and inhalation exposure (**Figure 3-4** and **Table 3-4**). The GM is slightly lower for inhalation exposure, but the difference to oral exposure is not significant (95% CI of the GM oral: 1.48 – 1.86, inhalation: 1.20 – 1.72). The 95% CI of the 75% percentile also has a large overlap (oral: 2.00 – 3.23, inhalation: 1.97 – 3.09). Consequently, the influence of exposure route for this comparison can be considered rather low.

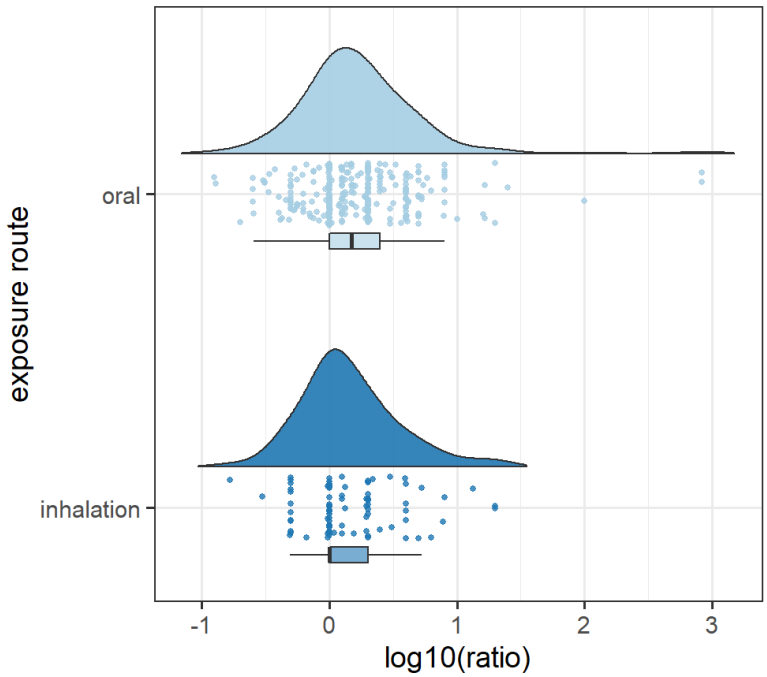


Figure 3-4 Distribution of ratios for the comparison of subacute with subchronic exposure, separated by exposure route.

Table 3-4 Summary statistics of distributions of ratios for the comparison of subacute with subchronic, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	8.15	67.67	1.65	2.78	0.42	1.50	2.50	7.98	303
inhalation	2.30	3.36	1.44	2.38	0.50	1.01	2.00	7.35	87

3.1.2.2 Influence of species

Analysis of the species influence on the ratios for the comparison subacute/subchronic revealed that mice have slightly higher ratios than rats (**Figure 3-5**). The arithmetic mean and standard deviation for rats are influenced by two extreme values (effects of Zearalenone on bodyweight in male and female rats), but the majority of parameters indicate higher ratios for mice (**Table 3-5**). Considering the spread of the data, these differences are likely not of relevance. In addition, the 95% CI of GM (rats: 1.34 – 1.74, mice: 1.48 – 2.01) and 75% percentile rats: 2.00 – 2.67, mice: 2.00 – 4.00) show a large overlap. Consequently, the species was not further evaluated as a possible contributing factor when other experimental factors were explored.

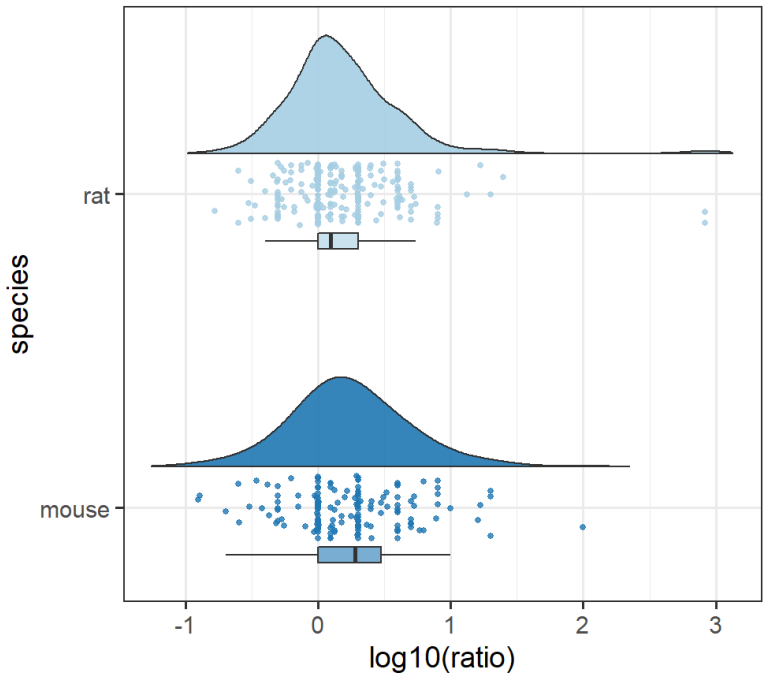


Figure 3-5 Distribution of ratios for the comparison of subacute with subchronic exposure, separated by species.

Table 3-5 Summary statistics of distributions of ratios for the comparison of subacute with subchronic, separated by species

Species	Mean	SD	GM	GSD	5%	Median	75%	95%	n
rat	9.24	76.89	1.52	2.70	0.50	1.25	2.00	5.38	233
mouse	3.29	8.49	1.73	2.68	0.41	1.90	3.00	8.00	157

3.1.2.3 Influence of sex

Sex of the test species does not seem to be an influencing factor on the ratios for the subacute/subchronic comparison (**Figure 3-6**). Again, extreme values by a single substance (Zearalenone) distort the arithmetic parameters and the percentiles and logarithmised location parameters are better suited for comparison with other distributions (**Table 3-6**).

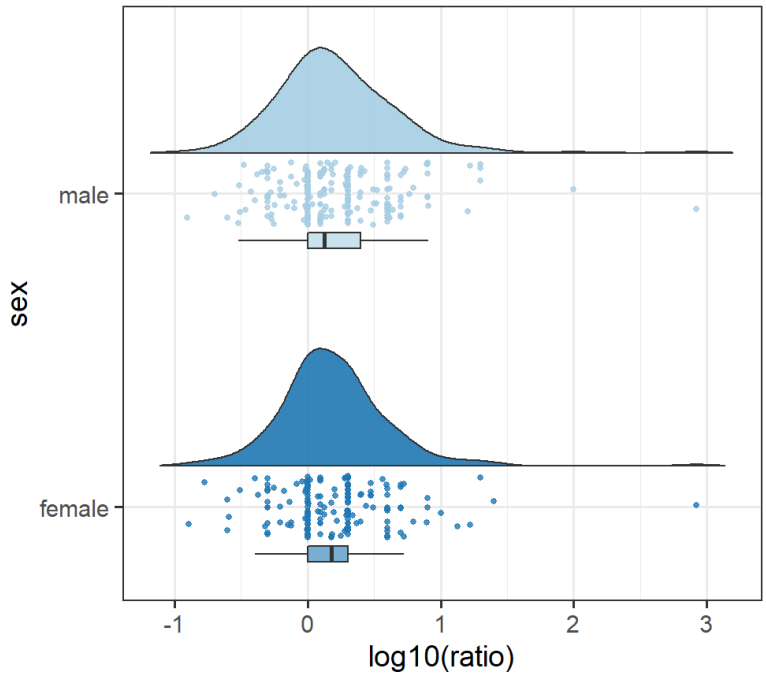


Figure 3-6 Distribution of ratios for the comparison of subacute with subchronic exposure, separated by sex.

The GM and median of the two distributions are relatively close together and the 95% CI of GM are overlapping to a large extent (male: 1.42 – 1.86, GM female: 1.37 – 1.84). The distributions are also not different regarding the 75% percentile as evidenced by a large overlap of the 95% CI (male: 2.00 – 4.00, female: 2.00 – 3.00). In conclusion, according to the analysed data, the sex of the test animals has no influence on the relevant parameters of the ratio distributions.

Table 3-6 Summary statistics of distributions of ratios for the comparison of subacute with subchronic, separated by sex

Sex	Mean	SD	GM	GSD	5%	Median	75%	95%	n
male	6.60	56.59	1.62	2.74	0.45	1.33	2.50	8.00	219
female	7.16	63.63	1.58	2.64	0.49	1.50	2.01	7.06	171

3.1.3 Subchronic/chronic

3.1.3.1 Influence of exposure route

At first glance the exposure route seems to have an influence on the resulting ratio distributions (**Figure 3-7** and **Table 3-7**). The GM is slightly lower for inhalation exposures than for oral. However, the 95% CI of the GM shows some overlap (oral: 2.86 – 3.27, inhalation: 2.27 – 3.00). Neither are the distributions different regarding the 75% percentile (95% CI oral: 4.76 – 6.00, inhalation: 4.00 – 6.46). Thus, although the route seems to have an influence on the lower percentiles, for the more relevant parameters this is not the case and the exposure route can be considered as not having an influence on the ratio distributions.

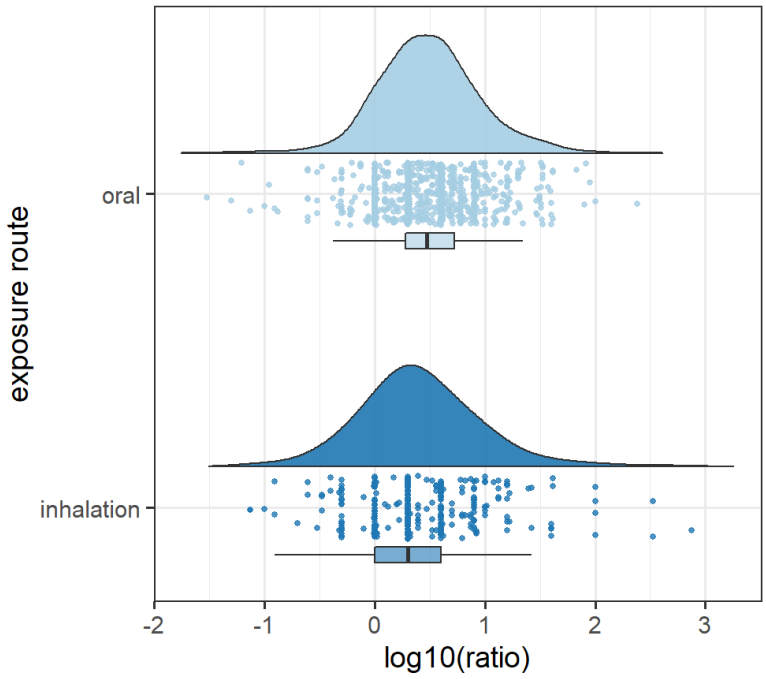


Figure 3-7 Distribution of ratios for the comparison of subchronic with chronic exposure, separated by exposure route.

Table 3-7 Summary statistics of distributions of ratios for the comparison of subchronic with chronic, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	5.7	11.75	3.06	2.85	0.64	3.00	5.33	19.12	895
inhalation	9.9	50.23	2.60	3.54	0.50	2.00	4.01	16.67	323

3.1.3.2 Influence of species

For the subchronic/chronic comparison, species seems not to be an influencing factor on the GM (95% CI rat: 2.61 – 3.12, mouse: 2.76 – 3.32). Neither for the 75% percentile a difference was observed (95% CI rat: 4.17 – 6.00, mouse: 4.67 – 6.67). The differences of other parameters are also rather small (**Figure 3-8** and **Table 3-8**), therefore the influence of species was not further considered when other factors were investigated.

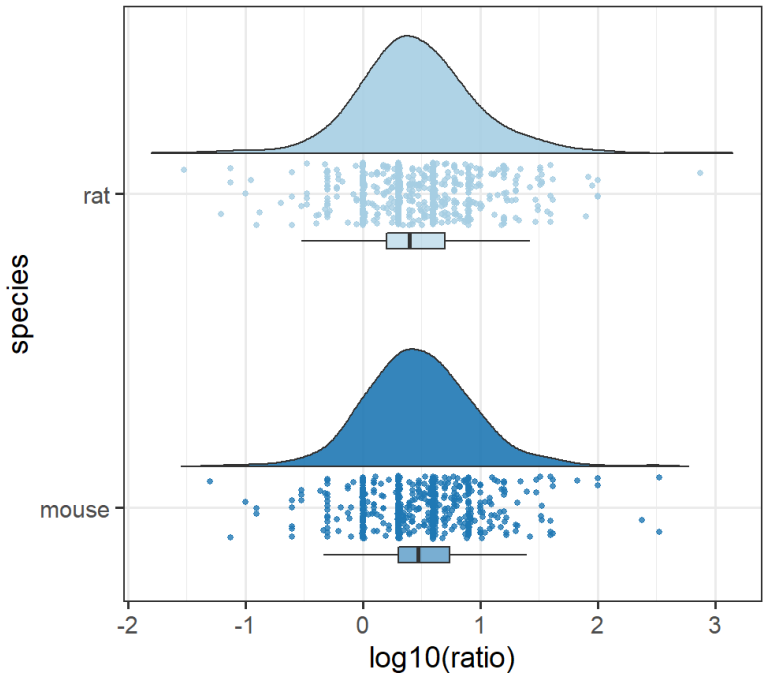


Figure 3-8 Distribution of ratios for the comparison of subchronic with chronic exposure, separated by species.

Table 3-8 Summary statistics of distributions of ratios for the comparison of subchronic with chronic, separated by species

Species	Mean	SD	GM	GSD	5%	Median	75%	95%	n
rat	6.76	31.01	2.85	3.09	0.50	2.50	5.00	20.00	653
mouse	6.88	23.56	3.03	2.98	0.50	2.96	5.50	16.58	565

3.1.3.3 Influence of sex

For the subchronic/chronic comparison the sex of the test animals seems to have a stronger impact than for the other time comparisons (**Figure 3-9**). The distribution for females has higher values across all parameters shown in **Table 3-9**. The 95% CI for the GM (male: 2.36 – 2.82; GM female: 3.06 – 3.68) does not show an overlap, indicating a high likelihood that this parameter is indeed dependent on the sex of the animals. Even for the 75% percentile a difference was determined, according to the 95% CI (male: 3.00 – 5.00, female: 5.00 – 7.67). Because the sex was not influencing the distributions of the other time comparisons and a differentiation by sex would be hard to implement in practice, the data is not analysed separately by sex in the downstream analysis.

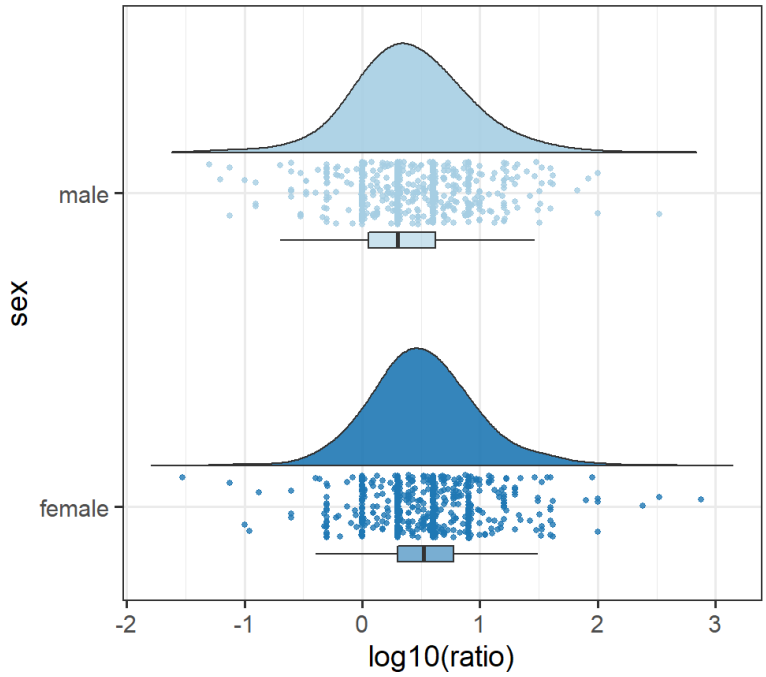


Figure 3-9 Distribution of ratios for the comparison of subchronic with chronic exposure, separated by sex.

Table 3-9 Summary statistics of distributions of ratios for the comparison of subchronic with chronic, separated by sex

Sex	Mean	SD	GM	GSD	5%	Median	75%	95%	n
male	5.36	15.42	2.57	3.04	0.48	2.08	4.17	16.27	639
female	10.07	60.60	3.34	3.10	0.50	3.57	6.00	20.95	597

3.1.4 Stratification by toxicity endpoints and route

The type of toxicity endpoint (effects on bodyweight, local or systemic toxicity) might also be an experimental factor with influence on the ratios. As inhalation exposure led to smaller ratios than oral exposure for all time comparisons (which however did not reach statistical significance for all comparisons or parameters), the analysis of the influence of the toxicity endpoint was performed separately for ratios derived from oral and inhalation studies.
Endpoint-specific ratios can, among the time comparisons, only be investigated for the subchronic/chronic comparison, as for subacute studies only effects on bodyweight were documented. Further, for oral studies the local effects were not documented, which is why a distribution of ratios derived from local effects can only be analysed for the inhalation route (**Figure 3-10**).

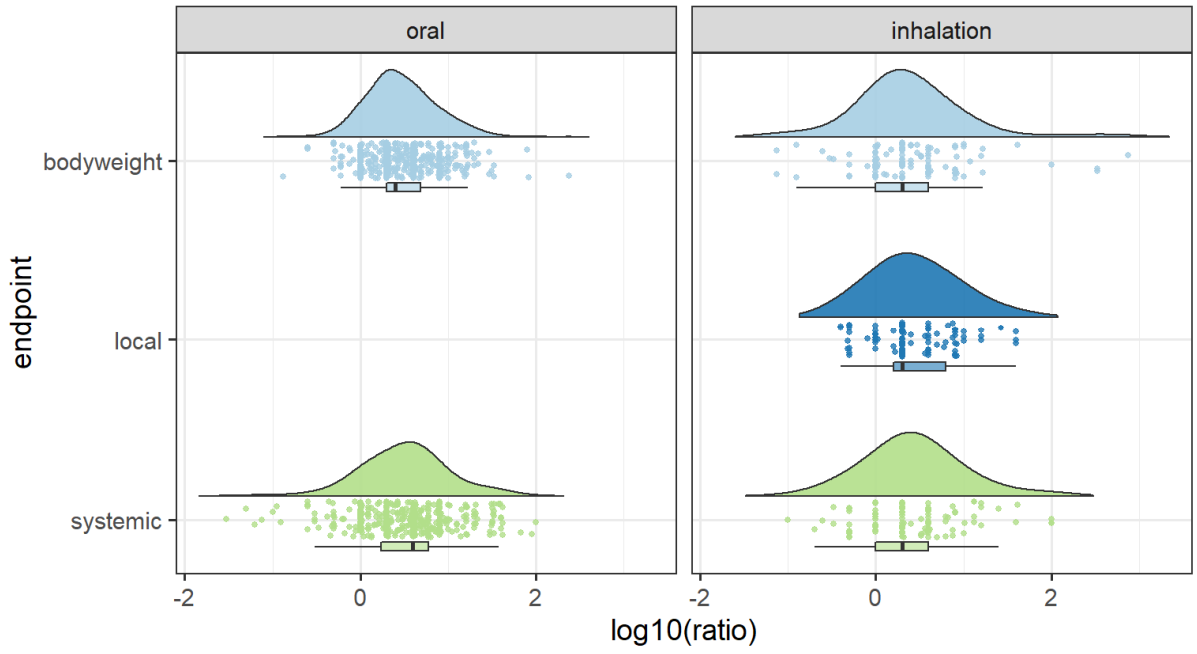


Figure 3-10 Distribution of ratios for the comparison of subchronic with chronic exposure, separated by exposure route and endpoint

Ratios from oral exposure are slightly higher for systemic effects than for effects on bodyweight when subchronic and chronic studies are compared (**Table 3-10**). In

relation to the overall spread of the distribution, this is likely of low practical relevance. Additionally, the 95% CI for the geometric mean (bodyweight: 2.72 – 3.24; systemic: 2.84 – 3.52) as well as the 75% percentile (bodyweight: 4.00 – 6.00; systemic: 5.00 – 7.67) are overlapping.

For inhalation exposure, the ratios from bodyweight and systemic effects are hardly distinguishable. The ratios of local effects are slightly shifted towards larger values, although overall these ratios are also quite comparable to bodyweight and systemic effects. The 95% CI of the local effects (geometric mean: 2.20 – 3.43; 75% percentile: 4.00 – 8.00) are largely overlapping with the systemic (geometric mean: 2.17 – 3.48; 75% percentile: 4.00 – 7.50) and bodyweight effects (geometric mean: 1.83 – 3.14; 75% percentile: 3.00 – 7.98).

Table 3-10 Summary statistics of distributions of ratios for the comparison of subchronic with chronic, separated by route and endpoint

Route	Endpoint	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	**bodyweight**	5.22	13.19	2.97	2.52	1.00	2.50	4.82	16.00	447
oral	**systemic**	6.18	10.10	3.17	3.19	0.50	3.99	6.00	27.00	448
inhalation	**bodyweight**	16.58	82.04	2.40	4.12	0.32	2.00	4.00	16.67	115
inhalation	**local**	5.17	7.71	2.73	2.96	0.50	2.00	6.25	16.03	101
inhalation	**systemic**	7.19	17.13	2.70	3.52	0.50	2.00	4.01	22.30	107

3.1.5 Stratification according to target organs (by route)

We further investigated whether differences in the ratio distributions can be observed for specific target organs of systemic toxicity. To this end, the ratios for effects on the two most frequently affected organs, liver and kidney, were compared to ratios for the remaining targeted organs. Besides liver and kidney, no further target organs were considered to be sufficiently represented in the dataset to be of value for the analysis. If a ratio was derived from effect levels where both organs (kidney and liver) were affected (6 ratios for the subchronic vs chronic comparison), this ratio was counted towards both groups.

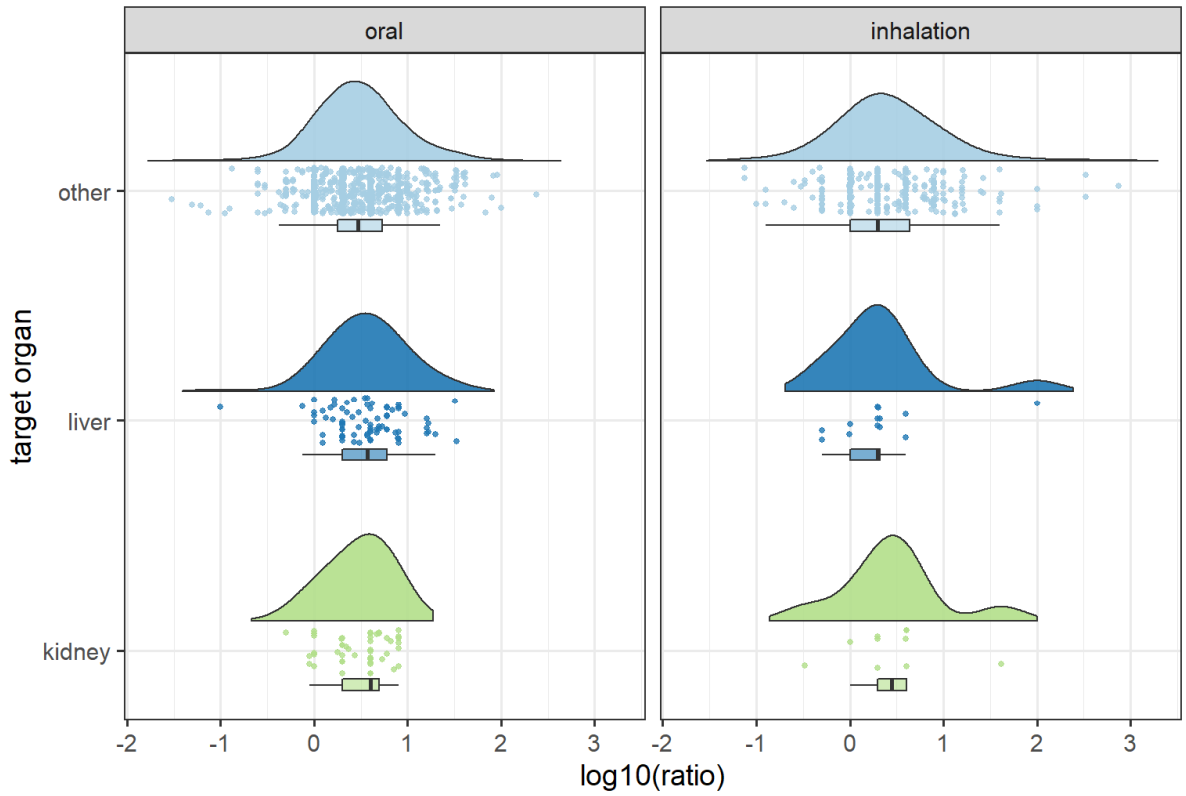


Figure 3-11 Distribution of ratios for the comparison of subchronic with chronic exposure, separated by exposure route and target organs. Only substances affecting the liver or kidney were treated as a separate group.

As systemic effects were not documented for subacute NTP studies, this analysis is only feasible for the ratios corresponding to subchronic-chronic extrapolation. As the exposure route might have an influence on ratio distributions (see e.g. section 3.1.1.1), the analysis was performed separately for oral and inhalation exposure (**Figure 3-11**). For inhalation exposure, the number of liver- or kidney-specific ratios is too small to allow a meaningful comparison. For example, the GM of ratios derived from inhalation studies with liver effects seems to be rather small (GM: 2.20), however the 95% CI covers a broad range (1.24 – 4.73). For oral exposure more data is available and 77 ratios for effects on the liver and 48 ratios for effects in the kidney could be derived. These are compared to 772 ratios which were derived from effects on other target organs or ratios which can't be clearly attributed to an effect on a specific organ (because the liver or kidney was only targeted at the LOAEL in one of the two compared studies).

The geometric mean of liver-specific ratios (GM 3.66) is greater than for the bulk of ratios for oral exposure (**Table 3-11**). At the 95% confidence level, this difference is not significant (liver-specific effects: 2.97 – 4.50, effects attributed to other organs: 2.81 – 3.26). The situation is the same at higher percentiles (the 95%CI of the 75% percentile are for liver 4.80 – 8.07, for other organs: 4.37 – 6.02).

Table 3-11 Summary statistics of distributions of ratios for the comparison of subchronic with chronic, separated by exposure route and target organs

Route	Target organ	Mean	SD	GM	GSD	5%	Me-dian	75%	95%	n
oral	other	5.84	12.48	3.02	2.93	0.60	2.96	5.33	20.00	772
oral	liver	5.55	6.12	3.66	2.54	1.00	3.75	6.00	16.67	77
oral	kidney	3.73	2.32	2.97	2.09	0.93	4.00	5.00	8.00	48
inhalation	other	10.03	51.79	2.61	3.55	0.50	2.00	4.40	16.67	300
inhalation	liver	9.48	27.22	2.20	3.75	0.50	2.00	2.10	42.40	13
inhalation	kidney	6.50	12.43	2.79	3.44	0.63	2.99	4.01	24.72	10

3.1.6 Stratification according to chemical features (by route and endpoint)

To investigate whether certain classes of chemicals may have an influence on the ratio distribution, ratios from experiments with chemicals belonging to two selected classes of chemicals were compared to the overall ratio distributions. The selection of investigated chemical classes was mainly driven by the number of members in the datasets after a manual screening. The two groups which were deemed suitable for a meaningful comparison are metal compounds (excluding metals in their elemental form) and alkylated aromatics (additionally containing benzene) (**Table 3-12**).

Table 3-12 Chemical groups (and their members) used in the analysis

Substance	CAS No.	Group
Divinylbenzene	1321-74-0	alkylated-aromatics
Methylstyrene	98-83-9	alkylated-aromatics
Cumene	98-82-8	alkylated-aromatics
Ethylbenzene	100-41-4	alkylated-aromatics
Vinyl Toluene	25013-15-4	alkylated-aromatics
Toluene	108-88-3	alkylated-aromatics
Benzene	71-43-2	alkylated-aromatics
Antimony Trioxide	1309-64-4	metal-compounds
Sodium chlorate	7775-09-9	metal-compounds

Substance	CAS No.	Group
Vanadium Pentoxide	1314-62-1	metal-compounds
Indium Phosphide	22398-80-7	metal-compounds
Sodium Nitrite	126-98-7	metal-compounds
Gallium Arsenide	1303-00-0	metal-compounds
Cobalt Sulfate Heptahydrate	10026-24-1	metal-compounds
Molybdenum Trioxide	1313-27-5	metal-compounds
Nickel Sulfate	10101-97-0	metal-compounds
Nickel Subsulfide	12035-72-2	metal-compounds
Nickel(II)Oxide	1313-99-1	metal-compounds
Manganese sulfate	10034-96-5	metal-compounds
Mercuric chloride	7487-94-7	metal-compounds
Stannous chloride	7772-99-8	metal-compounds
Sodium Dichromate Dihydrate	7789-12-0	metal-compounds

3.1.6.1 Influence of exposure route

Looking at individual time comparisons, there seem to be no differences in the ratio distributions between the groups of chemicals after oral exposure (**Figure 3-12**, **Figure 3-13**, **Figure 3-14** and **Table 3-13**, **Table 3-14**, **Table 3-15**). After inhalation exposure two differences merit a closer look: the alkylated aromatics after inhalation exposure appear to have a ratio distribution located at smaller values than the other groups of chemicals when the ratio is computed between subchronic and chronic or subacute and chronic studies (**Figure 3-12**, **Figure 3-14** and **Table 3-13**, **Table 3-15**). However, this observation does not hold for the other time comparison. The ratios for effects by alkylated aromatics even show a trend towards bigger ratios for the subacute/subchronic, although the number of ratios is rather low. Even though these differences between the different time comparisons could be mechanistically justified (e.g. the timeframe of subchronic studies is not sufficient for exacerbation of low dose effects by metal compounds), the analysis does not convincingly support the use of substance class-specific ratio distributions.

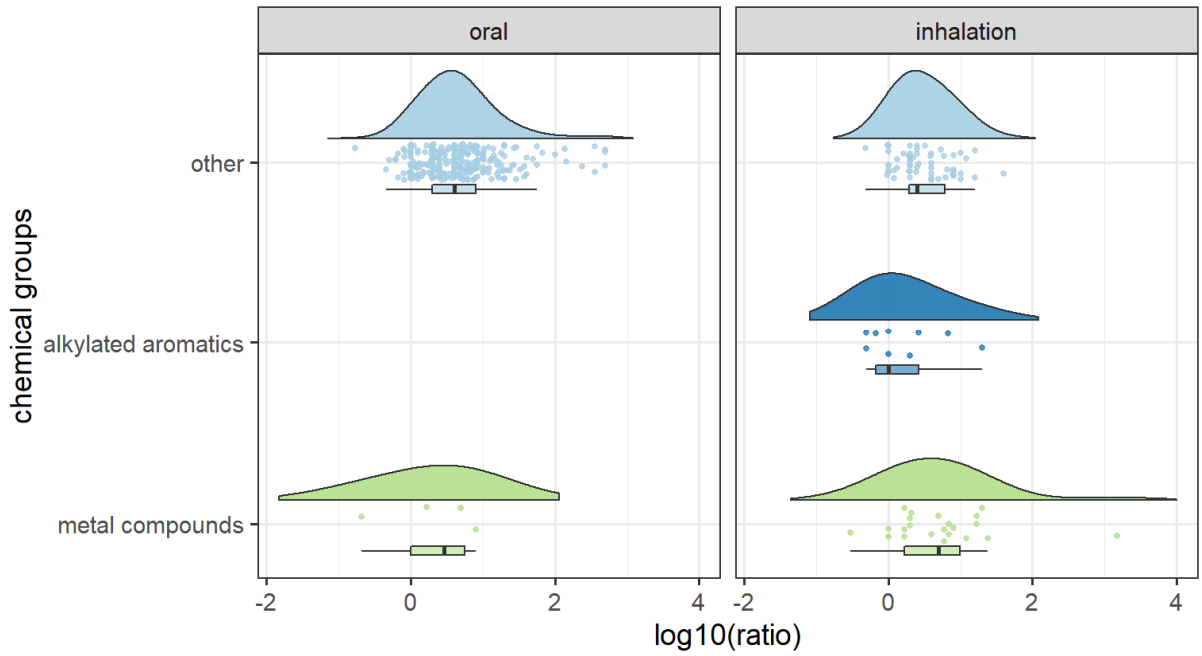


Figure 3-12 Distribution of ratios for the comparison of subacute with chronic, separated by chemical group and route

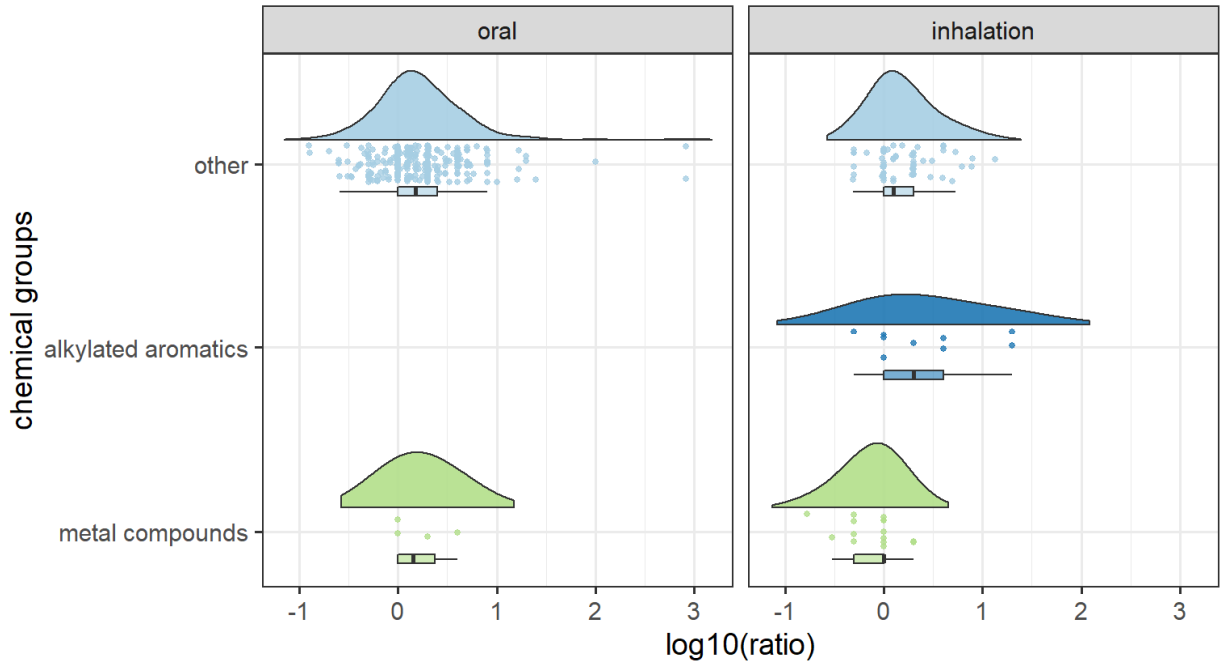


Figure 3-13 Distribution of ratios for the comparison of subacute with subchronic, separated by chemical group and route

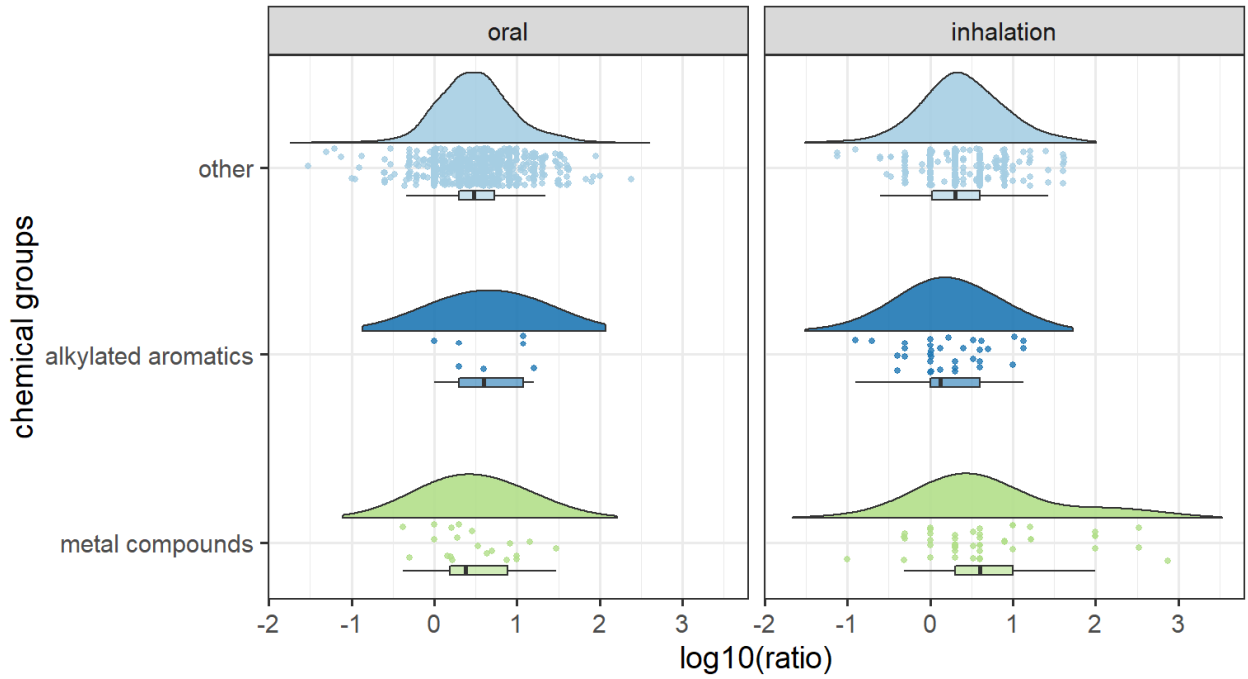


Figure 3-14 Distribution of ratios for the comparison of subchronic with chronic, separated by chemical group and route

Table 3-13 Summary statistics of distributions of ratios for the comparison of subacute with chronic, separated by chemical group and route

Route	Chemical group	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	**other**	15.66	59.81	4.42	3.35	1.00	4.00	8.00	33.00	301
oral	**metal compounds**	3.72	3.49	1.93	5.07	0.43	3.33	5.75	7.55	4
inhalation	**other**	4.72	5.91	3.06	2.43	1.00	2.50	6.12	12.40	59
inhalation	**alkylated aromatics**	3.89	6.35	1.70	3.51	0.50	1.00	2.67	14.67	9
inhalation	**metal compounds**	71.60	311.45	4.91	5.42	1.00	5.00	10.00	23.60	23

Table 3-14 Summary statistics of distributions of ratios for the comparison of subacute with subchronic, separated by chemical group and route

Route	Chemical group	Mean	SD	GM	GSD	5%	Me-dian	75%	95%	n
oral	**other**	8.23	68.12	1.65	2.79	0.42	1.50	2.50	7.98	299
oral	**metal compounds**	2.00	1.41	1.68	1.94	1.00	1.50	2.50	3.70	4
inhalation	**other**	2.11	2.18	1.54	2.10	0.50	1.25	2.00	6.16	63
inhalation	**alkylated aromatics**	5.94	8.07	2.65	3.79	0.70	2.00	4.00	20.00	9
inhalation	**metal compounds**	0.90	0.54	0.75	1.95	0.26	1.00	1.00	2.00	15

Table 3-15 Summary statistics of distributions of ratios for the comparison of subchronic with chronic, separated by chemical group and route

Route	Chemical group	Mean	SD	GM	GSD	5%	Me-dian	75%	95%	n
oral	**other**	5.70	11.87	3.06	2.85	0.65	3.00	5.33	19.56	868
oral	**alkylated aromatics**	7.00	6.14	4.49	2.97	1.30	4.00	12.00	14.80	7
oral	**metal compounds**	5.44	6.96	2.97	3.11	0.50	2.43	7.71	15.07	20
inhalation	**other**	4.58	7.03	2.47	2.95	0.50	2.00	4.01	16.00	245
inhalation	**alkylated aromatics**	2.98	3.61	1.63	3.16	0.32	1.33	4.00	11.58	33
inhalation	**metal compounds**	43.96	129.6	4.87	6.83	0.48	4.00	10.00	286.7	45

3.1.6.2 Influence of endpoint

Among the three time comparisons, substance-specific differences in the ratio distributions appear to be most prominent for the subchronic to chronic comparison of inhalation studies. This could possibly be due to the contribution of ratios from systemic (only available for the subchronic to chronic comparison) or local effects (only available for the inhalation path). Looking at the type of toxicity endpoint as an additional factor shows that this is probably not the case (**Figure 3-15** and **Table 3-16**). With the exception of local effects after exposure to metal compounds, there is no case where the ratio distribution based on effects on bodyweight deviates substantially from the distribution of local and systemic effects. Accordingly, the 95% CI of the GM shows a large overlap (e.g. alkylated aromatics, local effects: 0.48 – 1.91, alkylated aromatics, systemic effects: 1.79 – 5.51, alkylated aromatics, bodyweight: 0.93 – 2.37). For metal compounds however, the GM of ratios from local effects (1.73, 95% CI: 0.94 - 3.22) is significantly smaller than the ratio derived from bodyweight effects (8.32, 95% CI 4.53 – 16.65), but not from systemic effects (3.18, 95% CI 1.71 – 6.09).

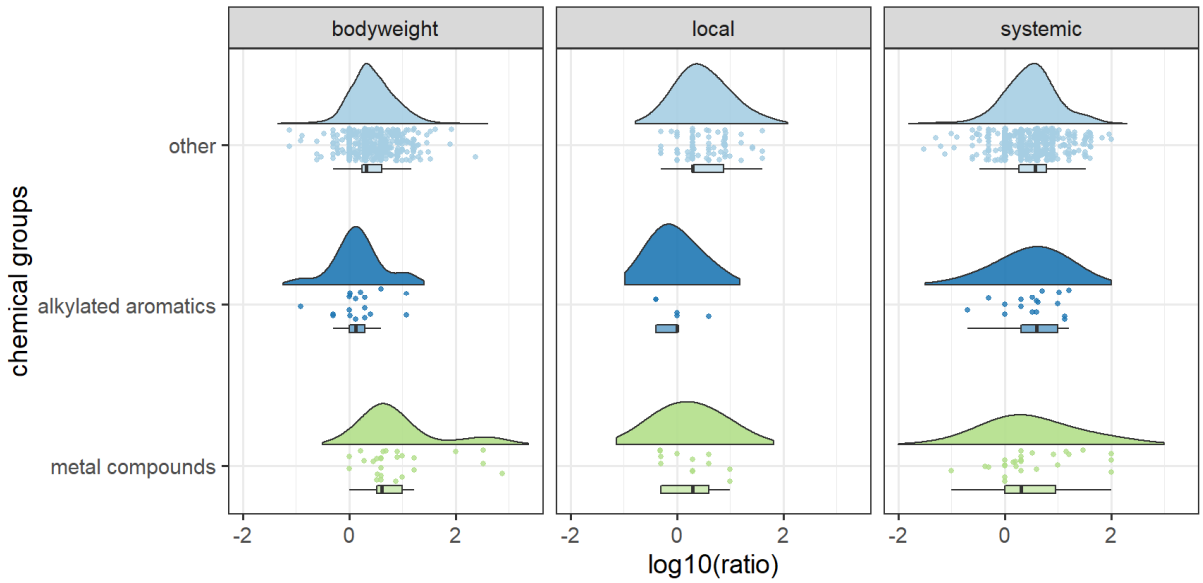


Figure 3-15 Distribution of ratios for the comparison of subchronic with chronic, separated by chemical group and endpoint

Table 3-16 Summary statistics of distributions of ratios for the comparison of subchronic with chronic, separated by chemical group and endpoint (oral and inhalation combined)

Endpoint	Chemical group	Mean	SD	GM	GSD	5%	Median	75%	95%	n
bodyweight	**other**	4.93	12.42	2.75	2.62	0.77	2.08	4.17	15.68	518
bodyweight	**alkylated aromatics**	2.62	3.52	1.48	2.92	0.44	1.33	2.00	12.00	18
bodyweight	**metal compounds**	63.13	166.4	8.32	5.68	1.17	4.16	10.00	333.3	26
local	**other**	5.70	8.25	3.11	2.83	0.55	2.00	7.50	25.07	84
local	**alkylated aromatics**	1.36	1.51	0.91	2.57	0.40	1.00	1.00	3.40	5
local	**metal compounds**	3.08	3.47	1.73	3.13	0.48	2.00	4.00	10.00	12
systemic	**other**	5.94	9.78	3.06	3.14	0.50	3.75	6.00	21.13	511
systemic	**alkylated aromatics**	5.51	5.07	3.22	3.40	0.44	4.00	10.00	13.87	17
systemic	**metal compounds**	15.13	31.26	3.18	5.58	0.45	2.00	9.17	100.0	27

3.2 Evaluation REACH data

3.2.1 Subacute/chronic

3.2.1.1 Influence of exposure route

The number of computable ratios for the comparison of subacute with chronic study results is rather low in the REACH dataset (**Figure 3-16** and **Table 3-17**). From the distributions of the available data, no differences between oral and inhalation exposure can be seen regarding the GM (95% CI for oral exposure: 1.42 – 3.96, for inhalation: 1.55 – 5.24). Also, the 95% CI of the 75% percentile shows a large overlap (oral: 3.63 – 7.67, inhalation: 3.20 – 10.67). In consequence, there is no indication to subset the dataset by route for further comparisons.

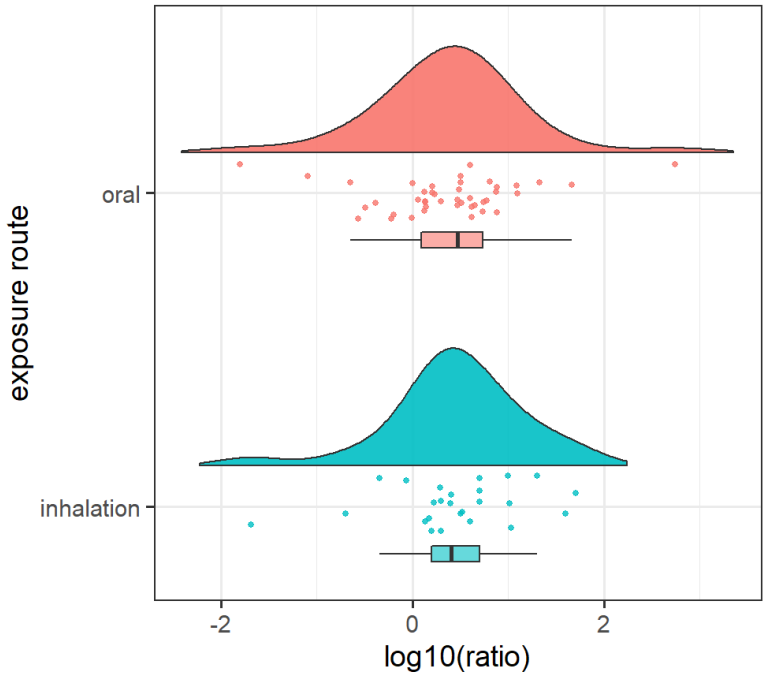


Figure 3-16 Distribution of ratios for the comparison of subacute with chronic, separated by exposure route

Table 3-17 Summary statistics of distributions of ratios for the comparison of subacute with chronic, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	17.39	83.51	2.35	5.40	0.23	2.96	5.48	20.53	43
inhalation	7.46	12.42	2.76	5.08	0.25	2.54	5.02	36.00	25

3.2.1.2 Influence of species

Comparisons of subacute with chronic study results are available from 3 species in the REACH dataset: rat (53 ratios), mouse (13), dog (2). The two values from the subgroup of dogs are included in the summary table (**Table 3-18**) but were omitted in **Figure 3-17** and not further considered for this assessment.

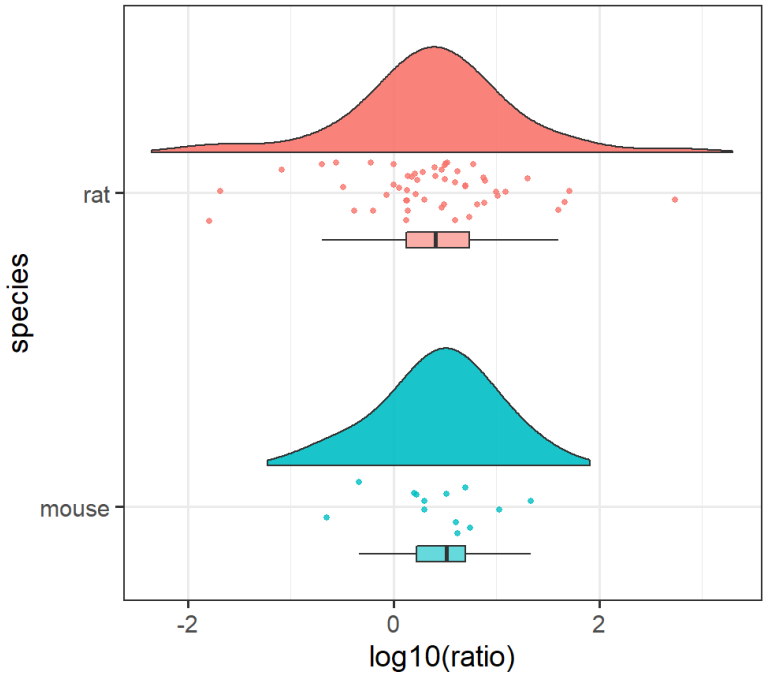


Figure 3-17 Distribution of ratios for the comparison of subacute with chronic, separated by species. Ratios computed from studies on dogs are omitted (only 2 values)

No difference in the GM between studies on rats and mice is evident in the data (95% CI of the GM for rat: 1.45 – 3.74, for mouse: 1.39 – 4.99). The point estimates for the 75% percentile are very close together and statistically not different form each other (95% CI for rat: 1.46 – 3.70, for mouse: 1.43 – 5.18).

Table 3-18 Summary statistics of distributions of ratios for the comparison of subacute with chronic, separated by species

Species	Mean	SD	GM	GSD	5%	Median	75%	95%	n
rat	16.13	75.49	2.35	5.86	0.15	2.54	5.41	42.55	53
mouse	4.77	5.70	2.69	3.34	0.36	3.25	5.02	14.97	13
dog	8.45	5.61	7.46	2.05	4.88	8.45	10.43	12.02	2

3.2.2 Subacute/subchronic

3.2.2.1 Influence of exposure route

The exposure route has no influence on the relevant parameters of the distribution (**Figure 3-18** and **Table 3-19**). The point estimates of both the GM and the 75% percentile are close together. Statistical testing indicates no difference between these parameters, the 95%CI for the GM is 1.12 – 1.39 for oral exposure and 1.07 - 2.05 for inhalation exposure. The same CI for the 75% percentile range from 2.00 – 3.00 (oral) and 2.00 – 4.77 (inhalation).

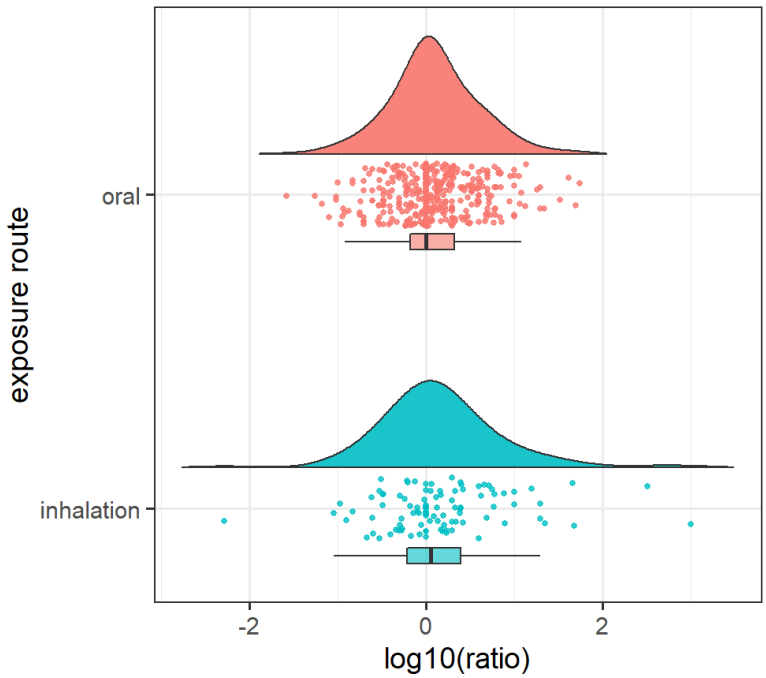


Figure 3-18 Distribution of ratios for the comparison of subacute with subchronic, separated by exposure route

Table 3-19 Summary statistics of distributions of ratios for the comparison of subacute with subchronic, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	2.58	5.33	1.24	3.10	0.20	1.02	2.12	8.95	379
inhalation	16.83	105.62	1.45	5.01	0.21	1.15	2.49	20.22	99

3.2.2.2 Influence of species

Splitting the dataset by species appears to result in a distribution shifted towards higher values for mice (**Figure 3-19**). However, this difference is not significant at the 95% confidence level of the GM (95% CI for rats: 1.12 – 1.41 and for mice: 1.12 – 2.40). The CI for 75% percentile also shows a large overlap and ranges from 2.00 – 2.71 for rats and from 2.00 – 5.71 for mice. In addition, the dataset contains 4 ratios computed from studies with dogs, which were omitted in the figure, but included in the table (**Table 3-20**).

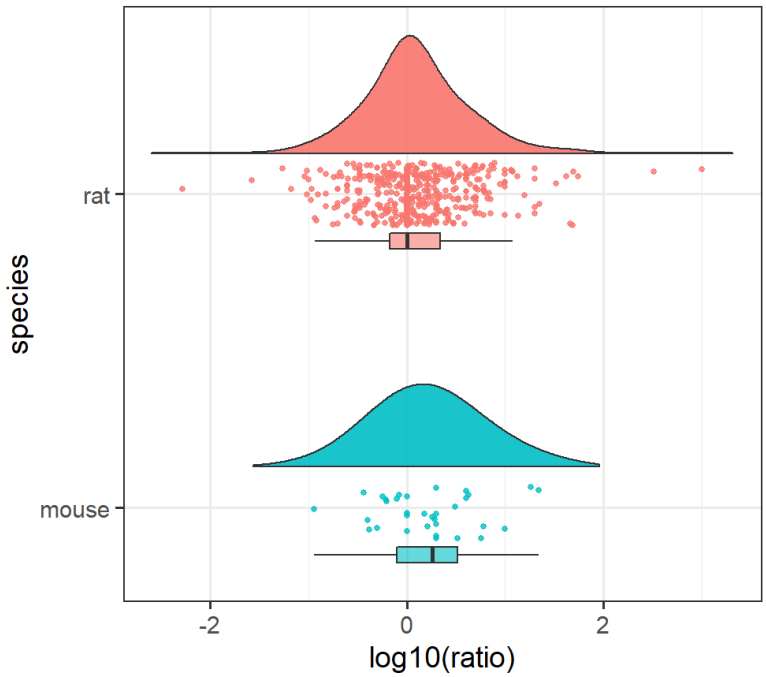


Figure 3-19 Distribution of ratios for the comparison of subacute with subchronic, separated by species. Ratios computed from studies on dogs are omitted (only 4 values)

Table 3-20 Summary statistics of distributions of ratios for the comparison of subacute with subchronic, separated by species

Species	Mean	SD	GM	GSD	5%	Median	75%	95%	n
rat	5.71	50.43	1.25	3.47	0.20	1.00	2.21	10.00	441
mouse	3.24	4.87	1.64	3.16	0.39	1.82	3.30	13.33	33
dog	4.28	4.16	1.67	8.54	0.32	3.94	6.95	8.72	4

3.2.3 Subchronic/chronic

3.2.3.1 Influence of exposure route

When subchronic studies are compared to chronic studies, the exposure route has no influence on the relevant parameters of the distribution (**Figure 3-18** and **Table 3-21**). The number of ratios is rather low with 96 ratios for oral and 48 for ratios from inhalation studies. Ratios from oral exposures have a lower GM, yet the difference of the GM between the two distributions is not statistically significant at the 95% confidence level (CI for oral exposure: 1.52 - 2.38, for inhalation exposure: 1.54 – 3.57). The 75% percentile is also lower for oral exposure but again, the difference is not statistically significant (95% CI for oral exposure: 2.54 – 5.47 and for inhalation exposure: 2.59 – 10.05).

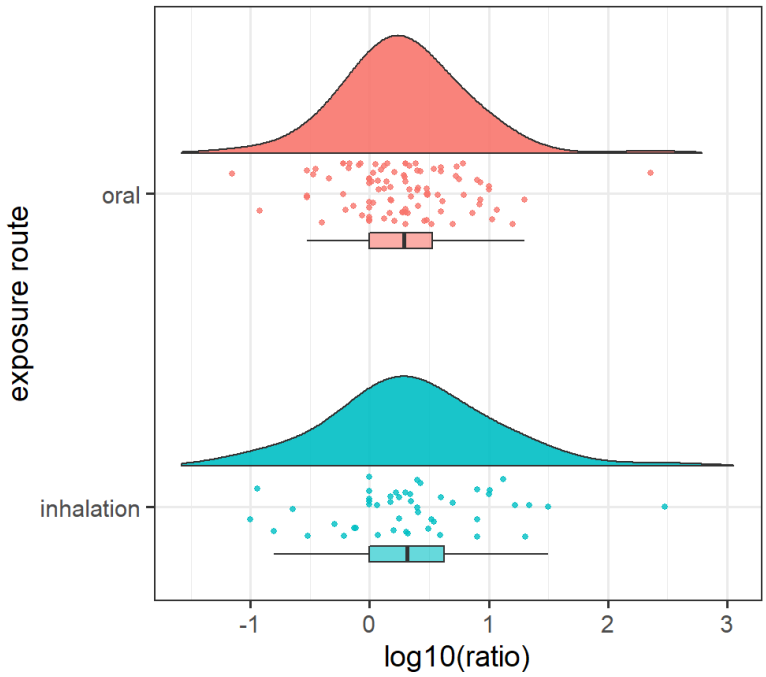


Figure 3-20 Distribution of ratios for the comparison of subchronic with chronic, separated by exposure route

Table 3-21 Summary statistics of distributions of ratios for the comparison of subchronic with chronic, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	5.35	23.49	1.91	3.12	0.33	1.95	3.36	10.19	96
inhalation	10.68	43.13	2.28	4.47	0.18	2.06	4.25	21.39	48

3.2.3.2 Influence of species

The dataset contains ratios for the comparison of subchronic and chronic studies in 5 species: rat, mouse, dog, primate and hamster. There was only one ratio each for primates and hamster, therefore these data were omitted in **Figure 3-21**. **Table 3-22** contains the data for all species. For the three species rats, mouse and dog, no differences in the distributions were identified. The differences in the GM (95% CI for in rats: 1.62 – 2.68; in mice: 1.50 – 3.96; in dogs: 0.65 – 2.35) and 75% percentile (95% CI for in rats: 2.55 – 5.58; in mice: 2.50 – 8.53; in dogs: 1.91 – 8.33) are too small to be significant in light of the variance of the data.

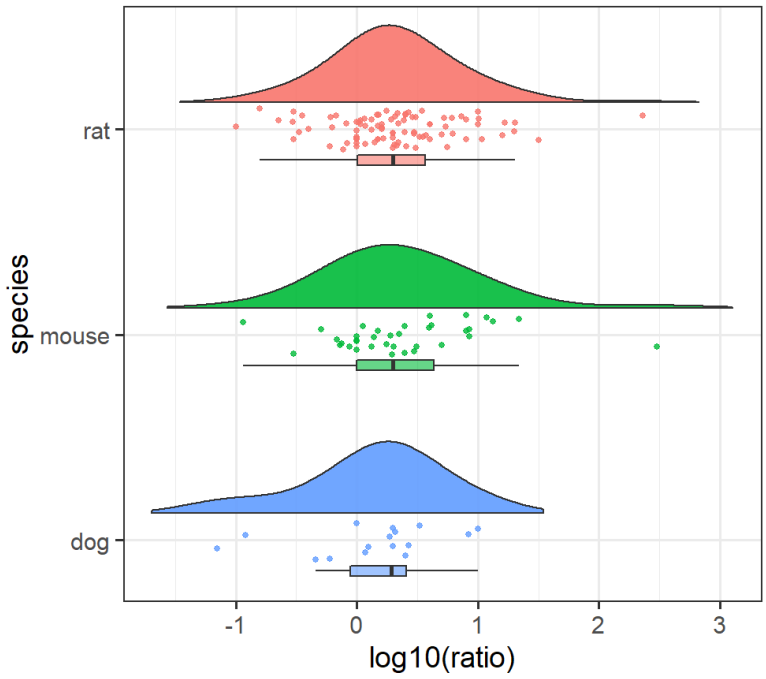


Figure 3-21 Distribution of ratios for the comparison of subchronic with chronic, separated by species. Ratios computed from studies in primates and hamsters are omitted (only 1 value each)

Table 3-22 Summary statistics of distributions of ratios for the comparison of subacute with subchronic, separated by species

Species	Mean	SD	GM	GSD	5%	Median	75%	95%	n
rat	6.15	24.47	2.08	3.35	0.30	2.00	3.69	16.33	90
mouse	11.98	49.58	2.40	4.00	0.46	1.99	4.34	15.47	36
dog	2.47	2.79	1.33	3.74	0.11	1.94	2.58	8.75	16
primate	0.60	-	0.60	-	0.60	0.60	0.60	0.60	1
hamster	1.00	-	1.00	-	1.00	1.00	1.00	1.00	1

3.2.4 Stratification according to target organs (by route)

The dataset derived from the REACH data contains only very few ratios which may be linked to liver or kidney specific toxicity (**Table 3-23**). It is not possible to draw any conclusions on target organ specific ratios from the REACH dataset.

Table 3-23 Summary statistics of distributions of ratios for the comparison of different study durations, separated by target organs.

Compari-son	Target organ	Mean	SD	GM	GSD	5%	Median	75%	95%	n
subacute/ chronic	**other**	13.76	67.21	2.43	5.23	0.21	2.95	5.21	34.43	67
subacute/ chronic	**liver**	12.42	-	12.42	-	12.42	12.42	12.42	12.42	1
subacute/ subchronic	**other**	5.58	48.81	1.28	3.50	0.20	1.00	2.36	10.00	471
subacute/ subchronic	**liver**	2.90	2.26	2.37	2.01	1.31	2.06	3.39	5.65	4
subacute/ subchronic	**kidney**	0.97	0.57	0.83	2.05	0.44	1.03	1.27	1.45	3
subchronic/ chronic	**other**	7.16	31.49	2.02	3.57	0.30	2.00	3.85	15.66	143
subchronic/ chronic	**liver**	2.00	-	2.00	-	2.00	2.00	2.00	2.00	1

3.3 Conclusion on stratification and relevance of experimental parameters

3.3.1 NTP data

Overall, the stratifications by experimental factors (route, species, sex, endpoint, target organs, structural properties of the test substances) revealed rather small contributions by the investigated factors.
There was a consistent trend over all comparisons that oral exposure leads to higher ratios than inhalation exposure when the parameters geometric mean and the 75% percentile are used to compare distributions. However, this was not statistically significant, and the effects were rather moderate in all cases.
The influence of the tested species also showed a consistent trend: study results on mice led to higher geometric means and 75% percentiles of the ratio distributions for all three time-comparisons. The numerical difference was smaller than the differences caused by the exposure route and did not reach statistical significance at the 95% level.
The sex of the animals only had an effect for the subchronic/chronic comparison. Females had a significantly higher geometric mean and 75% percentile of the ratio distributions than males and this difference was rather strong.

For the endpoint of toxicity (bodyweight, local, systemic) no consistent or statistically significant difference was identified.
Analysis of differences depending on target organs was only possible for ratios from oral studies, as there were not enough inhalation studies with kidney or liver effects to draw meaningful conclusions. Among the oral studies, the median of the ratios from substances which target the kidneys was significantly higher than the bulk of ratios from substances which were not targeting the liver or kidney. The geometric mean of the kidney-specific ratios as well as geometric mean and median of the liver specific ratios were also increased but did not reach statistical significance.

Finally, the two investigated substance classes, alkylated aromatics and metals, did not show a consistent influence on the ratio distributions. Both substance classes had rather low numbers of ratios.

3.3.2 REACH data

The reporting format of the REACH data made it significantly less likely to be successful in drawing conclusions on the influence of the toxicity endpoint and animal sex on the distributions. Therefore, these experimental factors were not used for modelling the distributions. Too few ratios that could be linked to kidney or liver toxicity were available to make any meaningful comparisons regarding ratios for organ-specific effects. Due to confidentiality issues, structural information of the substances linked to the derived ratios were not exported in the dataset used for modelling. This made it impossible to investigate the ratio distributions of substances with specific structural features.
The only experimental factors that could be investigated regarding their influence on the ratio distributions were the exposure route and the tested species. Both factors had no different effect on the geometric mean and 75% percentile of the derived distributions. Although their influence on other distribution parameters was not systematically analysed, the summary tables show no indication that this could be the case.

3.4 Proposal on distributions

The main purpose of the generated distributions is to compare them with currently used assessment factors (AF) in regulatory practice. For the latter, no distinction is made based on the experimental factors investigated here (e.g. toxicological endpoint or sex) and only a single AF is used for each time extrapolation. If an experimental parameter is in reality influencing the distribution but the used model is not taking into account this influence, performing time extrapolation using a single AF will carry a bias. This bias will be an under- or overestimation depending on the state of the parameter of the study which serves as POD and the bias will be greater if this state is underrepresented in the dataset that derived the single AF.

The GM of subchronic to chronic ratios from inhalation studies were slightly higher for local effects than for systemic effects. Also, the GM of ratios obtained for inhalation

data were slightly lower than those for oral data. But these differences were not statistically significant and not large enough to be considered meaningful. Accordingly, small differences in the GM or percentiles were observed for the investigated factors in most of the stratifications. However, the difference was statistically significant only in one occasion: the sex of the species when subchronic studies were compared with chronic studies and this effect was rather small. All the statistical comparisons were performed on basis of the geometric mean and 75% percentile. Although point estimates of other percentiles were not systematically checked, it is unlikely that they would present a different picture.

In conclusion, it seems most appropriate to use a single ratio distribution for each of the three time-comparisons and the two study sources. As discussed in section 4.2, the distributions derived from NTP data are considered more appropriate and reliable to serve as basis for AF (**Figure 3-22** and **Table 3-24**) and we propose to use the REACH data in a supportive manner. The ratios from the REACH data are presented in **Figure 3-23** and **Table 3-25**.

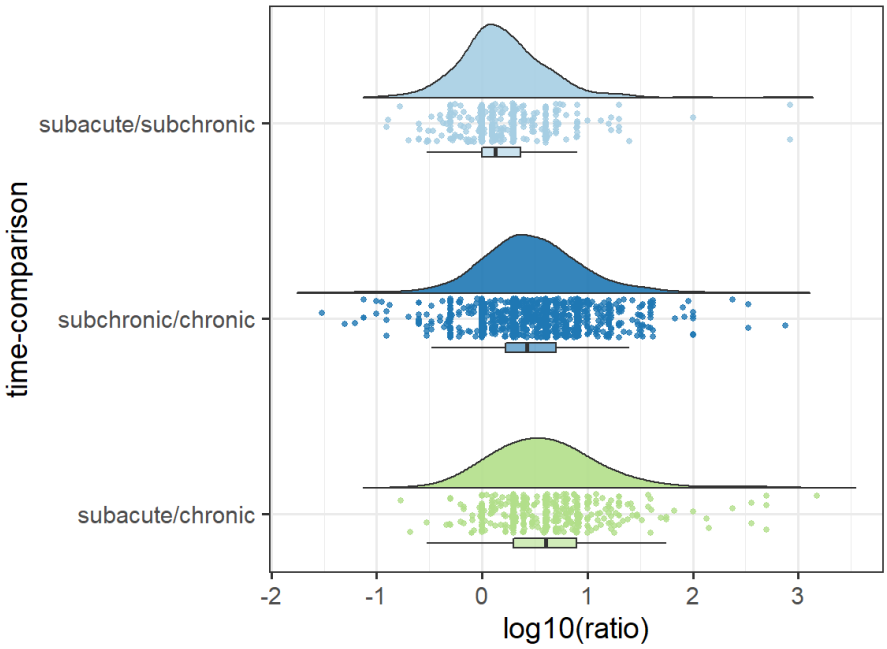


Figure 3-22 Distribution of all time comparison ratios derived from NTP data

Table 3-24 Recommended ratio distributions for each time extrapolation (based on NTP data evaluation).

Compari-son	Study source	Mean	SD	GM	GSD	5%	Median	75%	95%	n
subacute/ chronic	**NTP**	17.08	91.30	4.11	3.40	0.98	4.00	7.91	30.31	396
subacute/ subchronic	**NTP**	6.85	59.70	1.60	2.69	0.47	1.33	2.30	7.98	390
subchronic/ chronic	**NTP**	6.81	27.79	2.93	3.04	0.50	2.67	5.00	18.94	1218

Table 3-25 Ratio distributions derived from the REACH data for each time extrapolation.

Compari-son	Study source	Mean	SD	GM	GSD	5%	Median	75%	95%	n
subacute/ chronic	**REACH**	13.74	66.71	2.49	5.22	0.21	2.95	5.44	33.50	68
subacute/ subchronic	**REACH**	5.53	48.46	1.28	3.48	0.20	1.03	2.40	10.00	478
subchronic/ chronic	**REACH**	7.13	31.38	2.02	3.55	0.30	2.00	3.80	15.53	144

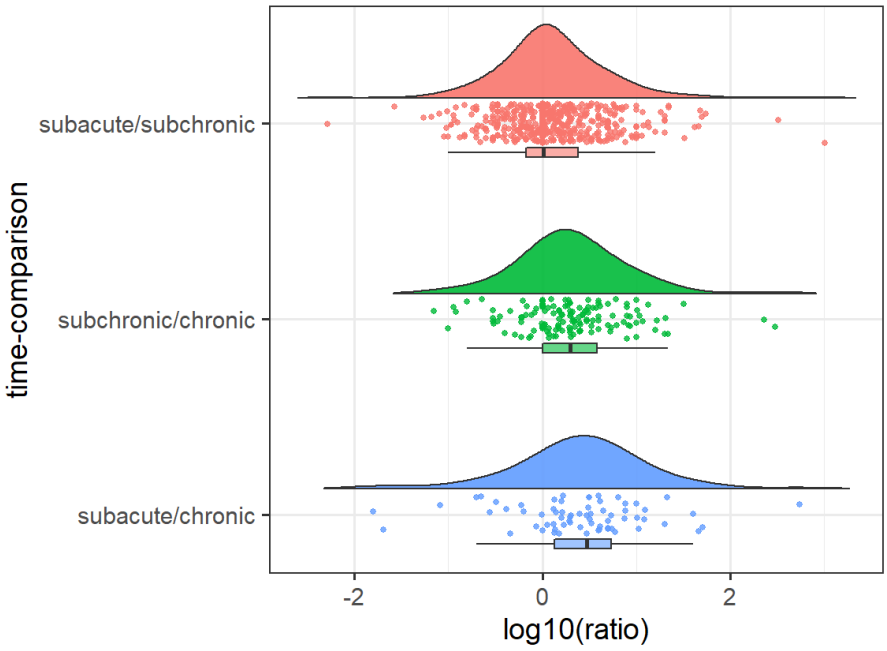


Figure 3-23 Distribution of all time comparison ratios derived from REACH data

4 Discussion

4.1 Methodological approach

Two different data sources were used to obtain the data for calculating the ratios which form the distributions, NTP technical reports and endpoint study records on repeated dose toxicity from REACH registration data. The two sources differ significantly in many aspects and different methodologies were used to derive the ratios. The impact of the differences on the resulting ratios are discussed below.

4.1.1 Amount and diversity of available study data

The REACH database contains over 20000 registered substances, many of which have data on repeated dose toxicity. In addition, the chemical spectrum covered by the substances in the REACH database is ideally suited for the purpose of studying time extrapolation factors.
However, the amount of study pairs that can be used to derive dose ratios is rather low, because often the selection criteria are not met. We intentionally used rather strict selection criteria, because under REACH much leeway is given to the registrants in how the studies are reported and also what they consider sufficient regarding quality of the study itself. Using less strict selection criteria would increase the number of derived ratios, but the resulting distributions are more likely to be affected by the varying quality of reporting and what individual registrants consider acceptable for fulfilling the information requirements.
NTP reports are available for much less substances compared to the REACH registrations. About 600 NTP reports are available. The spectrum of tested substances is mostly defined by industrial chemicals but is potentially biased towards suspected carcinogens, which might influence the derived distributions (Groeneveld et al. 2004).
In contrast to the REACH data, NTP reports cover at least two, often even three, different study durations under otherwise highly comparable experimental conditions. The quality of NTP studies is inherently high, so that no quality-related selection criteria were considered necessary. In consequence, relative to the amount of studies available, many more ratios can be computed from NTP reports than from REACH registrations.

4.1.2 Format and assessment of data

The data format is quite different between NTP and REACH data. The REACH data were extracted from the ECHA database and are structured according to the IUCLID data format, which follows the OECD harmonized templates. In principle this allows an efficient and systematic assessment of the data. Unfortunately, a great deal of relevant information is contained in natural language. This information was processed in a reproducible manner by coded rules. Generally, these rules were built iteratively after manual sighting the occurring entries and mistakes or misinterpretations are possible. The significance of this source of error is discussed in more detail below. The ECHA

database may alternatively be accessed via secondary data providers, e.g. EPA's CompTox Chemistry Dashboard or the OECD eChemPortal. To our knowledge, no secondary database is available which provides curated data using the information contained within natural language. Not using this additional information in natural language leads to a lower number of datasets that meet the selection criteria (e.g. a study was submitted with the species specified as “other:” (picklist value) with the additional specification in natural language as “rattus norvegicus”). An even more problematic consequence of not using all available information is the possible failure to detect data generated by read-across (Taylor et al. 2014).

The NTP data were manually assessed based on the technical reports published on NTP´s homepage. In contrast to the REACH data, where the effect levels are determined by the registrants, NOAELs and LOAELs were identified by ourselves, using the criteria outlined in the methods section. Manual assessment includes the possibility of random mistakes as well as certain degree of leeway. In this context, judgement of the adversity of histopathological changes is likely to be the major source of uncertainty. It was generally tried to follow the interpretation of the NTP study authors for the histopathological evaluations. For the effects on bodyweight, it was noted that the interpretation by the study authors is less consistent over time. Consequently, bodyweight was evaluated more rigidly based on the magnitude of effects. Great efforts were made to eliminate biases due to differences in endpoints investigated (e.g. organ weights or reproductive toxicity markers measured in the subchronic, but not chronic studies were not used for evaluation) in subchronic and chronic NTP studies (see section 2.1).

4.1.3 Study duration

The ratio distributions might be influenced by the way the duration of a study is determined and by the category boundaries for the compared time categories subacute, subchronic and chronic. Information from three variables (guideline, endpoint study records (ESR) category, study duration) was integrated to derive the correct study duration for the REACH data. This includes a manual screening for correctness of ambiguous cases, as described in detail in the methods section. Although mistakes at this point are still possible, we are confident that they are very scarce in the final dataset. Categorization of NTP studies into three different duration categories is trivial, under the premise of the following considerations for 14-day studies: The NTP 14-day study protocol was considered suitable for deriving subacute study results for the endpoint body weight. Batke et al. (2011) and Schröder et al. (2015) did not consider these studies in their analysis of subacute studies because of the different investigation depth compared to longer studies and the shorter study duration compared to the most frequent subacute study duration of 28 days. For the same reason, to avoid bias introduced by different investigation depths, we compared 14-day studies only regarding the clearly documented effects on bodyweight. For this endpoint the time difference between the 14-day and 28-day studies is likely not critical. For the REACH data, also the boundary definitions for subchronic and chronic study durations might potentially affect the results (see **Table 2-5**), however this is negligible

in practice, as the vast majority of study results follow the guideline durations (analysed, but not shown in detail).

4.1.4 Dose spacing

Zarn et al. (2011) and Zarn and O'Brien (2018) report slightly lower ratios when studies with a dose spacing > 5-fold were excluded. As the direction of this influence should depend on which of the two compared studies has the large dose spacing, other groups used the factor of the dose spacing between the two compared studies. Batke et al. (2011) got quite different ratio distributions depending on the factor of dose spacing difference but the sample sizes are too small to draw conclusions. Accordingly, in another analysis using the same definitions for dose spacing differences but a larger sample size, no differences were observed between the full and the reduced dataset (Lampe et al. 2018). Overall, it seems as dose spacing differences can have an influence on the ratios, yet most study pairs have a comparable dose spacing. In the NTP studies, the most common spacing factor is 2, and in a manual screening a spacing above a factor of $\sqrt{10}$ was rarely observed (apparently only in feeding studies with calculated doses from dietary uptake). For this reason the contribution of dose spacing to the results is probably minor and it was not evaluated as an influencing factor, as was also concluded by Takeshita et al. (2014).

4.1.5 Number of animals

Studies with a shorter duration usually have lower numbers of tested animals. Because of the lower statistical power, this potentially leads to higher effect levels for the shorter studies, which in turn might result in greater ratios. Zarn et al. (2011) compared the mean number of animals in the shorter study of study pairs that resulted in ratios <=1 with those that resulted in ratios > 1 for various time comparisons and found higher numbers of animals in the shorter study in those study pairs with ratios <= 1. Lampe et al. (2018) compared a subset of studies (subchronic studies with >= 10 animals) with the whole dataset und found no difference in the ratios. Effects by differences in animal number were not analysed in this report because little variation on animal numbers in the NTP and REACH studies is expected. It is also important to note, as holds true for other potential confounders, that effects due to different animal numbers do not lead to an over- or underestimation in terms of what the assessment factors are used for, i.e. the derived ratios reflect the typical study designs of the studies used as starting point for substance-specific assessments.

4.1.6 Study reliability (for REACH studies)

When Batke et al. (2011) compared the distributions from Klimisch reliability score (KS) 1 studies with the distributions form KS 2 studies, there was no difference in the location of the GM, but the ratios from KS 1 studies had a lower spread. Lampe et al. (2018) got the same result in their evaluation of the ECHA data, although with smaller differences. In our analysis, only studies from the REACH database with a Klimisch reliability score (KS) of 1 and 2 were selected for evaluation. We decided against

further restrictions based on KS as we noted that the provided metadata often does not fit the allocated KS and that the interpretation of KS criteria is different across study records in registrations. No attempt to harmonize or correct KS scores was made in this report, but the additional selection criteria (e.g. regarding the guideline conformity) likely removed illegitimate reliability classifications to some extent.

4.1.7 Influence of toxicity endpoints and target organs

No clear difference was observed in our study between endpoints body weight and (other) systemic endpoints for both oral and inhalation studies. The ratios subchronic/chronic for the endpoint local effects in inhalation studies was slightly increased compared to systemic effects, but without statistical significance. Still, it is noteworthy that the data do not point to lower assessment factors for local effects in the respiratory tract (local effects in oral studies were not analysed).

Further, there were no statistically significant differences between the ratios from substances with specific target organs compared to ratios derived from the remaining dataset. Only slightly increased GM values were observed for datasets with liver and kidney toxicity after oral exposure. No such differences were observed for inhalation studies (with much lower numbers of studies per target organ).

The evaluation by Batke et al. (2011) includes target organ specific AFs, but with a different definition than in this report: in their analysis, a target organ specific AF was the ratio between the lowest doses that affected the same organ in the compared studies, but the doses need to be above the general LOEL of the study pairs. The point estimates of the most relevant parameters of these target-organ specific distributions were not different from the AF for general effects, but the distribution had a smaller spread.

4.1.8 Influence of structural features

Due to confidentiality issues, no ratios linked to structural information could be exported from the isolated environment where the REACH data were processed. Accordingly, no analysis regarding structural features was performed with the REACH data. For the NTP data, the substances were manually allocated to two predefined classes (**Table 3-12**). No substantial differences between the two classes were found when compared to the overall database. A systematic approach, e.g. by molecular fingerprinting, might prove valuable to identify structural features influencing the ratio distributions.

4.1.9 Choice of dose descriptors and calculation of ratios

For each route, NOAELs or LOAELs were only used for comparisons when the units were equivalent (or could be converted to an equivalent unit as described in the methodology). For the NTP studies, this entailed the conversion of doses in feed or water to doses on a bodyweight basis. Bias introduced by conversion of doses is most

likely minor for time extrapolation, as doses from NTP studies are only compared within a study group where the same unit is used.
For deriving a ratio from REACH study data, calculation based on the two NOAELs was always preferred. If only LOAELs were available, these were used for calculation. The benefit of having a larger database for the ratio distributions (as opposed to a smaller database by only considering NOAEL ratios) probably outweighs the uncertainty introduced by adding LOAEL-derived ratios.
For the REACH data it was necessary to derive a single, concise value for the dose or concentration, as these are sometimes given as a range or unbounded value which cannot be directly used for calculation. The processing steps with probably the most potential impact on the final values are the handling of ranges and the handling of values specified by “>=” or “<=”. Values are given as a range in roughly 10% of cases. Our procedure selects the lower value of the range (except if that would be 0). A recent processing of the REACH study data also followed this approach (Saouter et al. 2018). Roughly 2% of the reported doses/concentrations are given with at least one qualifier as “>=” or “<=”. We handled these as “>” or “<” correspondingly with the consequence that some of these values are subsequently discarded (e.g. a NOAEL < 10 does not provide a useful information in the context of this analysis). Saouter et al. (2018) instead treated “>=”/”<=” equivalent to “=”. Although this seems justifiable from a mathematical standpoint, we believe that conversion to “>”/”<” better reflects the study outcome. In any case, the vast majority of the unbounded values are NOAELs with a “>=” qualifier, which are ultimately handled in the same way by both approaches. In addition, undefined doses or concentrations probably appear with a higher frequency in lower quality studies which play a smaller role in the ultimate data set used for deriving the ratios. Overall, the methodological choices of how to handle the dose/concentration value probably have no major impact on the ratios.

When deriving ratios from the REACH data, it is possible that several study results qualify as one of the two data points of the ratio (e.g. two NOAELs are available from two studies of the same duration in the same species). In this case, a two-tiered strategy was employed by first applying more stringent quality criteria (as described in the methods) and then calculating the mean in case there are multiple results left.
Approaches from other researchers to this issue are manifold but are usually limited to the second step by taking the minimum, mean or median value. We compared the use of mean and minimum for our data and did not notice a significant difference in the GM or median, but a slightly lower GSD if the mean is used. This fits to the assumption that calculating the mean would be less sensitive towards extreme values. Considering the rather small influence on the ratios this choice probably has only a minor effect on the resulting distributions. Further, it should be pointed out, that taking the mean instead of the minimum value at this point does not, by design, result in less conservative extrapolation factors, because this applies to numerator and denominator alike.

4.2 NTP versus REACH

The difference in the number of ratios which can be computed for each time extrapolation is a consequence of the difference in the data sources, the selection criteria and the rules for deriving a ratio from a study pair. These factors have already

been covered in the methodological discussion, but it is reiterated here that the high number of ratios from NTP studies is because the ratios were calculated separately for both sex and the three toxicological endpoints.
There are consistent differences between the distributions from REACH and NTP data. All distributions for time extrapolation from NTP data have a higher GM and median than the corresponding REACH distributions. The NTP distributions also consistently have a smaller GSD (**Table 4-1**). This means that for higher percentiles (roughly between 75% and 95%) the distributions from the two sources are quite comparable. At the 95% confidence level, these differences reach statistical significance for the subacute/subchronic and subchronic/chronic, but not for the subacute/chronic comparisons (**Table 4-1**).

Table 4-1 Summary of GM and median of the distributions derived by NTP and REACH data with the 95% CI for the GM and median (determined with 10000-fold bootstrap). CI which are not overlapping between the two study sources are highlighted in bold.

Comparison	**Study source**	**GM**	**95% CI (GM)**	**Median**	**95% CI (Median)**
subacute/chronic	NTP	4.08	3.61 – 4.62	4.00	3.31 – 4.02
subacute/chronic	REACH	2.49	1.67 – 3.72	2.95	1.70 – 4.00
subacute/subchronic	NTP	1.60	**1.45 – 1.76**	1.33	**1.25 – 1.62**
subacute/subchronic	REACH	1.28	**1.14 – 1.43**	1.03	**1.00 – 1.18**
subchronic/chronic	NTP	2.92	**2.73 – 3.11**	2.67	**2.50 – 3.24**
subchronic/chronic	REACH	2.02	**1.65 – 2.48**	2.00	**1.51 – 2.20**

An important observation is the large number of datasets with ratios below 1, which means that a lower NOAEL was observed in the shorter-term compared to the longer-term study. This was observed for both datasets, but to a much larger extent for the REACH data, especially for the comparison of subacute with subchronic studies (see e.g. Figure 3-18). The high frequency of low ratios < 1 has also been observed by others before (Lampe et al. 2018; Luechtefeld et al. 2016; Zarn et al. 2011). As discussed by Lampe et al. (2018), possible explanations comprise differences in dose spacing, adaptive processes and the influence of dose decrements. The fact that using benchmark dose-derived ratios does not change the high amount of ratios around 1 (Lampe et al. 2018) is an indicator that limitations in the study design which are at least partially compensated by BMD, like e.g. animal number, amount of dose groups or dose spacing, are not critical for the amount of ratios < 1.

We performed an in-depth analysis of some exemplary cases, prioritizing on especially low ratios (three examples shown in **Table 4-2** - **Table 4-4**). Our conclusion is that for the REACH data, a major cause for ratios below 1 is the inconsistent reporting of the data, e.g. effects which were decisive for the NOAEL in the shorter study were not

analysed (e.g. biochemical parameters) or not considered decisive (e.g. transient effects or adaptation) in the longer study. As evident from the detailed analysis of the exemplary cases, a manual assessment of the study reports would significantly alleviate the number of low ratios, but it is not feasible to do such an assessment on the scale of the whole dataset. In addition, as the full study reports are usually not available, often the reported effect levels remain doubtful, but no better alternative can be derived. The methodological decision to not differentiate between NOEL and NOAEL (and correspondingly for LO(A)EL) for the ratio calculation also contributes to the ratios < 1. This is a deliberate decision as we believe that many registrants used the terms “NOEL” and “NOAEL” interchangeable and that on the larger scale the benefits of this methodological decision outweigh the disadvantages. The following examples illustrate possible reasons for observed ratios <1 in the REACH database.

Table 4-2 First example of an in-depth analysis of a ratio well below 1 derived from the REACH database

Substance	Didecyldimethylammonium chloride		EC 230-525-2
Study duration	Effect levels available for comparison	Value used for comparison	Comments
Subacute	NOAEL: 2.5 mg/kg bw LOAEL: 27.5 mg/kg bw	2.5 mg/kg bw	Ratios based on NOAEL is given preference
Subchronic	NOAEL: 45.5 mg/kg bw LOAEL: 90 mg/kg bw	45.5 mg/kg bw	Ratios based on NOAEL is given preference
Derived ratio	2.5/45.5 = 0.0549		

Judgment after manual evaluation:

The NOAEL of the subacute study is based on clinical signs at the next higher concentration (27.5 mg/kg). The dose spacing is large between the NOAEL and LOAEL. The study is described as having doubtful reporting (without further details), yet it is a guideline study under GLP and the NOAEL seems reasonable.

The NOAEL and LOAEL of the subchronic study are justified.

It seems likely that the ratio of LOAELs would be more appropriate to use here, as the dose spacing in the subacute study probably leads to a particularly low value.

Derived ratio according to manual assessment: 27.5/90 = 0.306

Identified issues: study design

Table 4-3 Second example of an in-depth analysis of a ratio well below 1 derived from the REACH database

Substance	Butanone oxime		EC 202-496-6
Study duration	Effect levels available for comparison	Value used for comparison	Comments
Subacute	NOAEL: 4 mg/kg bw LOAEL: 20 mg/kg bw	4 mg/kg bw	Ratios based on NOAEL is given preference
Subchronic	NOAEL: 25 mg/kg bw NOAEL: 30 mg/kg bw NOAEL: 125 mg/kg bw LOAEL: 25 mg/kg bw LOAEL: 40 mg/kg bw	60 mg/kg bw (mean of 3 results)	Ratios based on NOAEL is given preference
Derived ratio	4/60 = 0.0667		

Judgment after manual evaluation:

The subacute NOAEL of 4 mg/kg bw is reported to be based on haematological effects at the LOAEL. In addition, changes in spleen weight are reported at the LOAEL. The animals recovered from most effects after 14 days. A second subacute study, reporting a LOAEL at the lowest tested dose of 250 mg/kg bw, is available in the REACH database but did not pass the selection criteria (for no readily apparent reason) or was not included in the database at the time of the database dump.

The NOAEL of the first subchronic study (25 mg/kg bw) is based on erythrotoxicity with matching histopathological findings. The NOAEL of 125 mg/kg bw in another subchronic study is based on neurobehavioral effects. For this study also a LOAEL for haematological effects at 40 mg/kg bw was reported, which was the lowest tested dose. The NOAEL of 30 mg/kg bw cannot be found anymore in the disseminated database.

The subacute NOAEL may be overly conservative and based on transient effects. Yet, the studies give a consistent picture as the hematopoietic system being the primary target irrespective of the exposure duration. The haematological effects at the subacute NOAEL can be considered the first signs of this mode of action and this NOAEL should be respected. Of the three subchronic NOAELs, only the value of 25 mg/kg bw passes the manual assessment.

Derived ratio according to manual assessment: 4/25 = 0.16

Identified issues: NOAEL reporting by registrants, choice of critical effect, updated data

Table 4-4 Third example of an in-depth analysis of a ratio well below 1 derived from the REACH database

Substance	Alkali salt of substituted aryl amino sulfonyl triazinyl amino sulfonyl aryl diazo aryl sulfonyl sulfonate		EC 413-090-5
Study duration	Effect level available for comparison	Value used for comparison	Comments
Subacute	NOEL: 24 mg/kg bw	24 mg/kg bw	NOEL is treated as NOAEL for ratio calculation
Subchronic	NOAEL: 250 mg/kg bw	250 mg/kg bw	
Derived ratio	24/250 = 0.096		

Judgment after manual evaluation:

The subacute NOEL is based on increased GPT (glutamic-pyruvic transaminase) activities in the next dose group. Because of the absence of histopathological liver lesions, this was not considered an adverse effect in the registration. A NOAEL was not reported, but likely is 600 mg/kg bw, the highest tested dose.

The subchronic NOAEL of 250 mg/kg bw is based on numerous effects at the next higher dose (1000 mg/kg bw). This NOAEL is justified.

No other study results are reported in the disseminated REACH database.

Derived ratio according to manual assessment: 600/250 = 2.4

Identified issues: Equivalence of NOEL and NOAEL for ratio calculation, NO(A)EL reporting by registrants

With the experiences made while preparing the REACH data, the NTP data is considered to provide the database with the higher quality, which is probably also reflected by the higher spread of the REACH derived distributions. The number of ratios derived for subacute/chronic and subchronic/chronic study pairs from the REACH data is also quite low, as relatively few chronic studies are reported in the registration dossiers. Therefore, the distributions derived from the NTP data are given preference.

4.3 Comparison with published data

Comparison with published data was based mainly on the GM, as all relevant studies on this topic reported the GM. The most relevant measure for the spread of the

distributions is the GSD[5], which is reported in most work, but not all. For the sake of comparison between studies, when no GSD was given, the GSD was estimated assuming a lognormal distribution based on all reported percentiles using the package 'rriskDistributions' (Belgorodski et al. 2017). Estimated GSD are indicated as dashed lines and may have a considerable error. If no GSD is indicated in **Figure 4-1** to **Figure 4-3**, a fit to a lognormal distribution was not possible.

Across different research groups, the chosen database and methodological choices vary considerably and looking at the data in **Figure 4-1** to **Figure 4-3** it is obvious that this is affecting the resulting distributions. The most prominent differences between our analysis in this report and the body of the available literature on this topic are discussed below.

4.3.1 Subacute/chronic

Reported GM for subacute/chronic comparisons range from 1.8 – 10, but most values lie between 3.2 (Lampe et al. 2018) and 4.9 (Groeneveld et al. 2004) (**Figure 4-1**). The results obtained from our evaluation of NTP studies are shown separately for inhalation and oral studies, to allow comparison with other route-specific evaluations. Although we observed slightly lower values for inhalation data, we consider this difference a random observation. This conclusion is supported by the sum of evaluations, which do not point to systematic differences between routes (**Figure 4-1**).

The two extreme values are both from distributions with a small sample size (Schröder et al. (2015): GM 1.8, 23 ratios, Kramer et al. (1995): GM 10.0, 10 ratios): The GM obtained by Schröder et al. (2015) of 1.8 is lower than those from all other evaluations. Interestingly, the median of this dataset is much higher (median=3.2). The authors state that the datasets evaluated are lognormally distributed, which should lead to GMs very close to the median. This difference between GM and median point to the existence of several values <<1. No explanation is given on the difference between GM and median.

Kramer et al. (1995) observed a very high GSD of 15 for the 10 ratios for subacute to chronic NOAELs from inhalation studies. The high variability might have been caused by individual outliers. Also, the authors allowed ratios being calculated from studies of different species (e.g. subacute study in rats and chronic study in mice), which is expected to increase the variability. Substantially lower values were obtained by the authors for the oral data, as presented in the publication by Kramer et al. (1996) (GM 4.1, GSD 4.4, n=71,); a slightly higher GM of 6.5 was documented in the underlying RIVM report for a set of 57 values from oral data (Kramer et al. 1995).

The study by Kalberlah et al. (2002) also used NTP study data, but differs in several aspects of the methodology used for the evaluation. They set all ratios which would

[5] GSD is used in many previous publications to characterise the spread of distributions. But note that it does not represent the standard deviation of the lognormally distributed data or is defined by statistical terms otherwise. GSD in the presented figures in this section is dimensionless and can only semi-quantitatively be used to compare the spread of two distributions.

have resulted in a ratio < 1 to 1. This obviously would shift the distribution to higher values, but the authors also provide an alternative analysis which includes ratios <1. This alternative distribution is not much different and still has a considerably higher GM than the corresponding distribution in this report. Additionally, and probably more critical, their analysis did not account for the different investigation depth of the 14-day studies compared to (sub)chronic studies. With the few endpoints analysed in the 14-day studies, this most likely increased the ratios between subacute and chronic studies. This problem was avoided in our evaluation by using only data on body weight for comparing 14-day study results with subchronic or chronic studies.

In perspective, the results from our analysis of NTP studies are in the middle of all compared point estimates for the GM. Our REACH data results (with a GM at the lower end of values) are included for comparison but are considered less reliable for reasons discussed above.

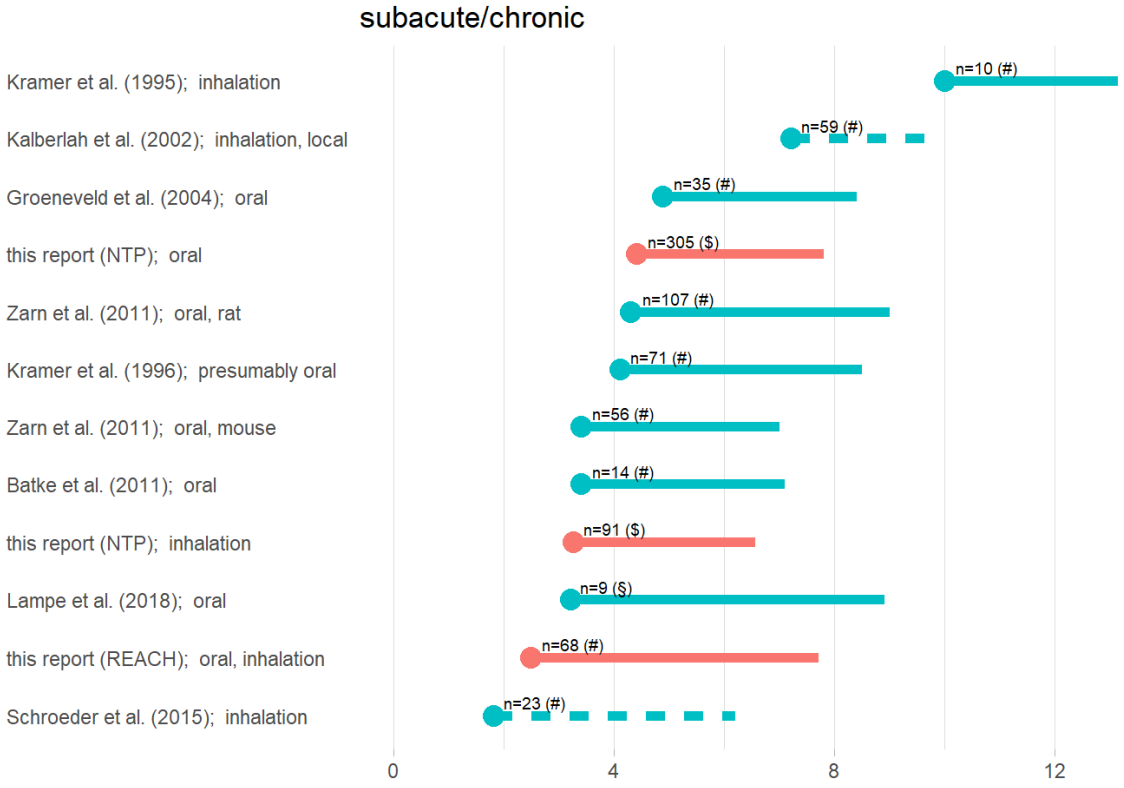


Figure 4-1 Published distributions of subacute/chronic effect level ratios. Shown are the reported GM (dot) and the dimensionless GSD (range from the GM to the end of each line) on the normal scale. When no GSD was reported, the GSD was estimated (dashed line). Lines extending beyond 12.5 were cut off. Selections of the underlying database (route and species, species is mentioned only if datasets were stratified for single species) are given after the literature reference. (#) = substance level ratios, i.e. one ratio per substance. (§) = ratios at study level, i.e. ratios may be derived from multiple studies per substance. ($) = separate ratios per sex, i.e. two ratios may be derived from two compared studies. Data from this report are highlighted in red.

Zarn et al. (2011) observed a slightly lower GM for mice than for rats. Such a trend is not visible in our data (see section 3.1.1.2). These authors also discussed that the dose decrement might have an effect on the ratios, as is discussed in detail by Zarn et al. (2011). The dose decrement describes the decrease of the effective dose over lifetime of the animals at a given food concentration, because of a decreasing food consumption per kg bodyweight with increasing age. When using average food consumption data, this might result in higher NOAELs for longer-term studies. Yet, this reflects the actual situation how chronic studies are evaluated and would therefore not lead to an overestimation of the time extrapolation factor as it is intended for current regulatory practice. Note that the evaluations of Zarn and O'Brien (2018) are not included here, as these are containing mainly the same data as in Zarn et al. (2011), but ratios were obviously calculated based on food concentrations.

For substances acting locally in the respiratory tract, some authors argue that these effects are concentration-driven and not depending on exposure time. With this argumentation, ECETOC proposed assessment factors of 1 for differences in study duration (ECETOC 2010). This was not confirmed by Schröder et al. (2015), who found higher values for local effects (eye and respiratory tract) for all comparisons (subacute to chronic, subacute to subchronic, subchronic to chronic). Also in our evaluation slightly higher values for local effects (although not statistically significant) appeared for the comparison subchronic to chronic (see section 4.3.3; not evaluated for subacute data).

Restricting the evaluation of subacute NTP studies to effects on body weights compensated for the large differences in investigation depth between the NTP 14-day and 2-year studies. However, for use in regulatory toxicology the difference between **any type of effect level** in both studies (e.g. the difference in NOAELs between an OECD TG 407 and an OECD TG 452 study) is more relevant and might be underestimated by an endpoint-specific approach, as the variability is assumed to be smaller (see also the discussion in section 4.3.3). But as the list of endpoints investigated in the NTP 14-day studies is smaller than that of the OECD 407 or 412 studies, this limitation is necessary and might be seen as a compensation of the relatively short exposure period of 14 days compared to the typical subacute study with 28 days.

4.3.2 Subacute/subchronic

Overall, the GM of published distributions in literature are quite homogeneous and range from 1.3 – 2.4. Only the value reported by Kalberlah et al. (2002) for inhalation data of locally acting substances is higher (GM=3.3), but as discussed above the lower investigation depth in the NTP 14-day studies is likely responsible for the difference between this value and our results. The GM of the NTP-derived distribution from our dataset is in the middle of this range. The corresponding GM for the REACH data is at the lower end of 1.3, just as the results by Lampe et al. (2018) and Batke et al. (2011).

The analysis by Lampe et al. (2018) used REACH data as well (amongst other data sources, subacute NTP data were excluded due to the 14 day duration) and is

comparable to our analysis of the REACH data regarding the selection criteria. The study by Batke et al. (2011) used data from the RepDose® database, which contains study data from several sources, including NTP. Subacute NTP data were however excluded as well because of the 14-day study duration. The study by Batke et al. (2011) also differs in other methodological aspects (inclusion of organ specific effect levels) and generally tends to derive small point estimates, not just for the subacute/subchronic comparison. This work was extended and updated by Escher et al. (2020), who included additional studies from REACH registrations and other sources. The new GM values reported by Escher et al. (2020) et al. are slightly higher than those given by Batke et al. (2011) and very close to the results of our evaluation (**Figure 4-2**).

Slightly higher GMs were reported by Kramer et al. (1995) (GM 2.2; again a high GSD of 7.4 was observed) and by Takeshita et al. (2014) (GM 2.4). The latter also tried to discriminate between various endpoints (effects on liver, kidney, blood, and body weight). They found a lower GM (0.53) for effects on blood, but not for the other endpoints. Due to the high variability observed and the low number of comparisons per endpoint (n=9 for blood effects) the reliability of these observations is doubtful. We did not observe meaningful differences between effects on liver or kidney compared to the remaining datasets when analysing subchronic and chronic studies in our much larger NTP study dataset (section 3.1.5).

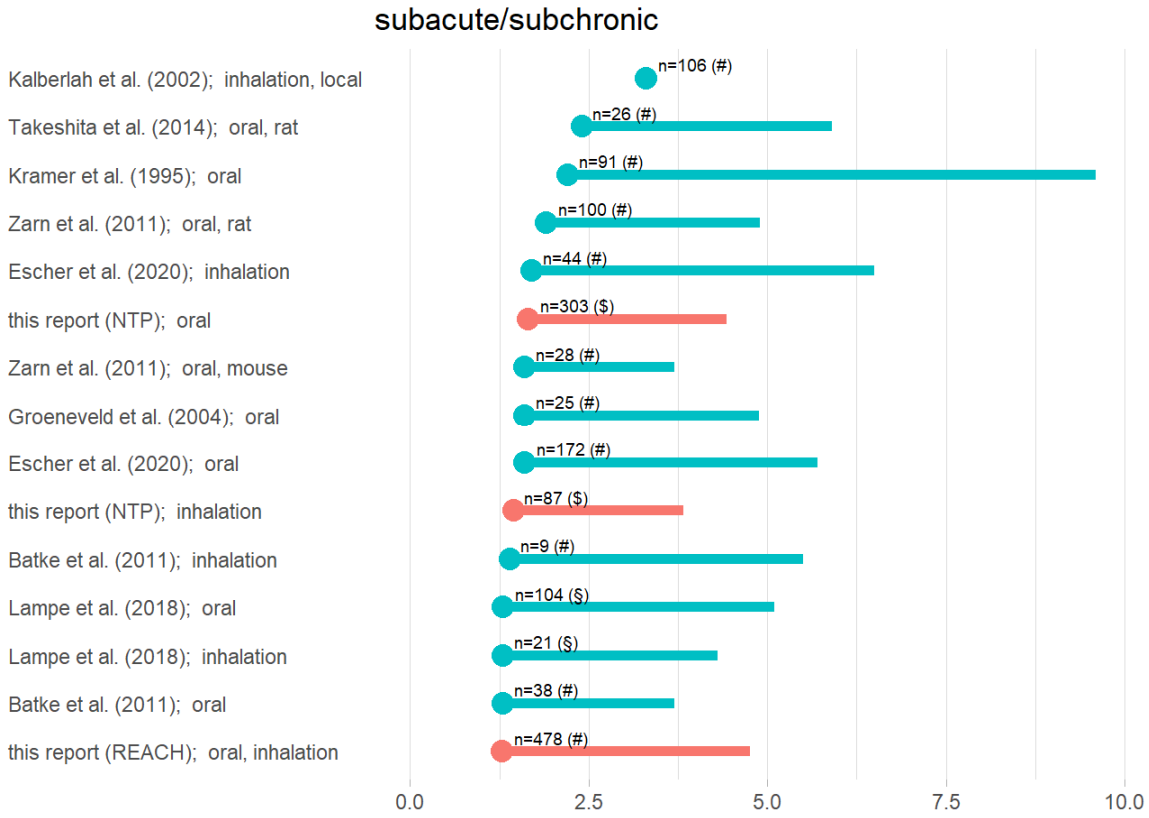


Figure 4-2 Published distributions of subacute/subchronic effect level ratios. Shown are the reported GM (dot) and the dimensionless GSD (range from the GM to the end of each line) on the normal scale. If no GSD is indicated, an estimation was not possible (see beginning of section 4.3). Selections of the underlying database (route and species, species is mentioned only if datasets were stratified for single species) are given after the literature reference. (#) = substance level ratios, i.e. one ratio per substance. (§) = ratios at study level, i.e. ratios may be derived from multiple studies per substance. ($) = separate ratios per sex, i.e. two ratios may be derived from two compared studies. Data from this report are highlighted in red.

4.3.3 Subchronic/chronic

GMs for the subchronic/chronic comparison range from 1.4 – 3.1. For this comparison of study durations, the result of our analysis represents the highest GM among all compared results. Here, the GM fits quite well to the methodologically comparable analyses by Kalberlah et al. (2002).

Data evaluations by Zarn et al. (2011) of studies with pesticides yielded ratios, which are slightly lower, but in a similar range (GMs 2.2 and 2.5 for oral mouse and rat data, resp.). However, evaluations done based on the RepDose® database (Batke et al. 2011; Escher et al. 2020) are reporting values (**Figure 4-3**), without obvious explanation for the differences observed. However, it is noted that the GMs of 1.3 for

subacute/subchronic ratios and 1.4 for subchronic/chronic ratios reported by Batke et al. (2011) do not fit to the GM for subacute-chronic ratios of 3.4 (the discrepancy is even more striking at the level of medians). In consequence, the ratios obtained for the sub-steps do not explain the overall differences observed by the same authors between subacute and chronic studies. No GM for subacute/chronic is reported by Escher et al. (2020).

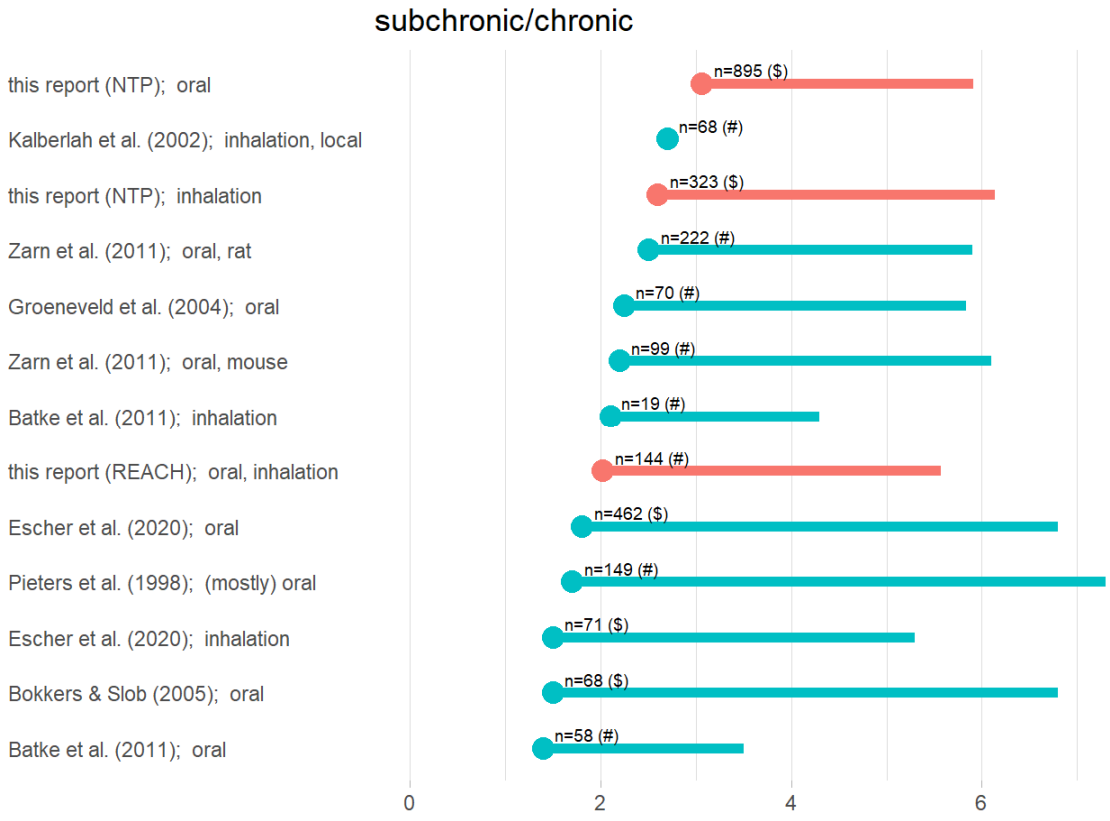


Figure 4-3 Published distributions of subchronic/chronic effect level ratios. Shown are the reported GM (dot) and the dimensionless GSD (range from the GM to the end of each line) on the normal scale. If no GSD is indicated, an estimation was not possible If no GSD is indicated, an estimation was not possible (see beginning of section 4.3). Selections of the underlying database (route and species, species is mentioned only if datasets were stratified for single species) are given after the literature reference. (#) = substance level ratios, i.e. one ratio per substance. (§) = ratios at study level, i.e. ratios may be derived from multiple studies per substance. ($) = ratios of endpoints, i.e. multiple ratios may be derived from two compared studies. Data from this report are highlighted in red.

A major disadvantage of evaluating NOAEL ratios is the dependency of NOAELs on the initially chosen experimental doses. In a major effort Bokkers and Slob (2005) analysed dose-response data from 31 NTP studies by dose-response modelling and

compared the obtained benchmark dose (BMD) ratios with NOAEL ratios. They obtained similar geometric means: the GM based on NOAEL ratios (reported in **Figure 4-3**) was 1.5. It increased slightly to 1.7, when BMD ratios were calculated, whereas GSD decreased from 5.3 to 2.9, indicating a lower variability for BMD ratios. In our evaluation of subchronic to chronic ratios GSD was 3.04, similar to the variability observed by Bokkers and Slob (2005) for the BMD ratios.

Bokkers and Slob (2005) evaluated a subset of the NTP studies in our evaluation and obtained substantially lower GMs (irrespective of whether BMD or NOAEL ratios were calculated) compared to our results. A possible explanation are methodological differences: Bokkers and Slob (2005) derived endpoint-specific ratios for the two endpoints body weight and liver weight. In contrast, our analysis aimed to identify the lowest NOAEL for each study type. Although the analysis of Bokkers and Slob (2005) provides insight into the development of a specific effect with prolonged exposure time, regulatory practice asks for a factor suitable to estimate the (equivalent) point of departure from an adequate long-term study.

Similar to Bokkers and Slob (2005), Lampe et al. (2018) derived benchmark dose ratios for subacute versus subchronic studies (see section 4.3.2) for a subset of studies for 20 chemicals. They found slightly lower values for BMD ratios compared to NOAEL ratios and slightly lower GSDs. However, (Lampe et al. 2018) did not find a major influence on BMD ratios, when they either calculated BMD ratios for the same endpoint or ratios from the lowest BMD per study (GM 1.1 in both cases). Interestingly, they found a high percentage of 45% for ratios below 1, for which, among other possibilities, they hold responsible the reversibility of effects over time.

5 Conclusions

We evaluated two large datasets: NTP studies and repeated dose studies from REACH registrations. Due to observations on reporting quality of REACH data and the restrictions of a largely automated study evaluation, we consider the REACH database less reliable than the manual evaluation of NTP data. This conclusion is supported by a larger GSD for REACH data, pointing to higher variability in this dataset.

The evaluation of both datasets led to the conclusion that no consistent differences with regard to time extrapolation for the variables

- Route of application (oral, inhalation)
- Sex
- Species
- Toxicity endpoints after inhalation (local, systemic)
- Target organs
- Substance classes (exemplary examined for two groups of substances with NTP data)

are evident.

Therefore, we conclude that the combined dataset of ratios from oral and inhalation NTP data is adequate for proposing distributions for time extrapolation. The data are derived from studies on 256 substances, from which close to 400 (subacute/subchronic and subacute/chronic each) or more than 1200 (subchronic/chronic) ratios were calculated (with separate evaluation of species, sex and selected endpoints, i.e. body weight, local and systemic effects). The same distributions as for systemic effects are proposed for local effects in the respiratory tract after inhalation, a conclusion, which is supported by publications pointing to similar or higher ratios for local compared to systemic effects.

This report presents one of the few data evaluations covering both sub-steps (subacute/subchronic and subchronic/chronic**)** as well as the full span (subacute – chronic) of the most frequently used time extrapolation steps. The following distributions were obtained from the NTP data evaluation.

Table 5-1 Recommended ratio distributions for each time extrapolation (based on NTP data evaluation).

Comparison	Mean	SD	GM	GSD	5%	Median	75%	95%	n
subacute/chronic	17.08	91.30	4.11	3.40	0.98	4.00	7.91	30.31	396
subacute/subchronic	6.85	59.70	1.60	2.69	0.47	1.33	2.30	7.98	390
subchronic/chronic	6.81	27.79	2.93	3.04	0.50	2.67	5.00	18.94	1218

The ratios obtained for subacute versus chronic and subacute versus subchronic studies fit very well to other evaluations published in recent years. The ratios for subchronic versus chronic studies are at the upper end of the range observed in recent evaluations. However, multiplication of GMs or medians of the two sub-steps yield

values in agreement with the subacute – chronic ratios, indicating consistency in these datasets.

References

Batke, M., Escher, S., Hoffmann-Doerr, S., Melber, C., Messinger, H., Mangelsdorf, I. (2011). Evaluation of time extrapolation factors based on the database RepDose. Toxicology Letters 205, 122-129.

Belgorodski, N., Greiner, M., Tolksdorf, K., Schueller, K., Flor, M., Göhring, L. (2017). Package 'rriskDistributions', Online: https://cran.r-project.org/web/packages/rriskDistributions/rriskDistributions.pdf (08.11. 2018).

Bokkers, B.G., Slob, W. (2005). A comparison of ratio distribution based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. Toxicological Sciences 85, 1033-1040.

ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals. (2010). Technical Report No. 110. Guidance on Assessment Factors to Derive a DNEL, Brussels, Belgium.

Escher, S.E., Mangelsdorf, I., Hoffmann-Doerr, S., Partosch, F., Karwath, A., Schroeder, K., Zapf, A., Batke, M. (2020). Time extrapolation in regulatory risk assessment: The impact of study differences on the extrapolation factors. Regulatory Toxicology and Pharmacology 112, 104584.

Groeneveld, C., Hakkert, B., Bos, P., Heer, C. (2004). Extrapolation for exposure duration in oral toxicity: a quantitative analysis of historical toxicity data. Human and Ecological Risk Assessment 10, 709-716.

Guha, R. (2007). Chemical Informatics Functionality in R. Journal of Statistical Software 18, 16pp.

Kalberlah, F., Föst, U., Schneider, K. (2002). Time extrapolation and interspecies extrapolation for locally acting substances in case of limited toxicological data. Annals of Occupational Hygiene 46, 175-185.

Kalberlah, F., Schneider, K. (1998). Quantification of Extrapolation Factors. Final Report of the Research Project No. 116 06 113 of the Federal Environmental Agency, Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund Fb 797 Wirtschaftsverlag NW Bremerhaven.

Kramer, H.J., van den Ham, W.A., Slob, W., Pieters, M.N. (1995). Derivation of Conversion Factors to Estimate an Indicative Chronic NOAEL from Short-Term Toxicity Data. Report No. 620110 001, RIVM Rijksinstituut voor Volksgezondheid en Milieu Bilthoven Netherlands.

Kramer, H.J., van den Ham, W.A., Slob, W., Pieters, M.N. (1996). Conversion factors estimating indicative chronic no-observed-adverse-effect levels from short-term toxicity data. Regulatory Toxicology and Pharmacology 23, 249-255.

Lampe, B.J., Fuller, E., Kuppusamy, S.P. (2018). A quantitative comparison of points of departure between 28-day and 90-day repeated dose studies with a proposed extrapolation factor. Regulatory Toxicology and Pharmacology 92, 189-200.

Luechtefeld, T., Maertens, A., Russo, D.P., Rovida, C., Zhu, H., Hartung, T. (2016). Analysis of public oral toxicity data from REACH registrations 2008-2014. ALTEX - Alternativen zu Tierexperimenten 33, 111-122.

Pieters, M.N., Kramer, H.J., Slob, W. (1998). Evaluation of the uncertainty factor for subchronic-to-chronic extrapolation: statistical analysis of toxicity data. Regulatory Toxicology and Pharmacology 27, 108-111.

Saouter, E., Biganzoli, F., Ceriani, L., Versteeg, D., Crenna, E., Zampori, L., Sala, S., Pant, R. (2018). Environmental Footprint: Update of Life Cycle Impact Assessment methods – Ecotoxicity freshwater, human toxicity cancer, and non-cancer. EUR 29495 EN, Publications Office of the European Union. https://ec.europa.eu/jrc/en/publication/environmental-footprint-update-life-cycle-impact-assessment-methods-ecotoxicity-freshwater-human.

Schröder, K., Escher, S., Licht, O., Simetska, N., Mangelsdorf, I. (2015). Time Extrapolation Factors for Local Effects in Inhalation Studies. Project number [40191]. Draft March 2015, Fraunhofer Institute for Toxicology and Experimental Medicine ITEM, Chemical Risk Assessment, Databases and Expert Systems. On behalf of Umweltbundesamt, Germany.

Takeshita, J., Kishimoto, A., Gamo, M. (2014). Investigation of organ-specific assessment factors related to sub-acute and sub-chronic toxicity studies. Fundamental Toxicological Sciences 1, 1-6.

Taylor, K., Andrew, D.J., Rego, L. (2014). The added value of the 90-day repeated dose oral toxicity test for industrial chemicals with a low (sub)acute toxicity profile in a high quality dataset. Regulatory Toxicology and Pharmacology 69, 320-332.

Zarn, J.A., Engeli, B.E., Schlatter, J.R. (2011). Study parameters influencing NOAEL and LOAEL in toxicity feeding studies for pesticides: exposure duration versus dose decrement, dose spacing, group size and chemical class. Regulatory Toxicology and Pharmacology 61, 243-250.

Zarn, J.A., O'Brien, C.D. (2018). Current pesticide dietary risk assessment in light of comparable animal study NOAELs after chronic and short-termed exposure durations. Archives of Toxicology 92, 157-167.

List of Tables

List of Figures

indicated, an estimation was not possible (see beginning of section 4.3).Selections of the underlying database (route and species, species is mentioned only if datasets were stratified for single species) are given after the literature reference. (#) = substance level ratios, i.e. one ratio per substance. (§) = ratios at study level, i.e. ratios may be derived from multiple studies per substance. ($) = ratios of endpoints, i.e. multiple ratios may be derived from two compared studies. Data from this report are highlighted in red. (Pieters et al. 1998) 70

REPORT 7:
Interspecies Extrapolation

RESEARCH PROJECT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Summary

Extrapolating from experimental animal studies to humans is a key step in deriving OELs. Existing methods either use default factors (often split in a toxicokinetic and a toxicodynamic part) and/or apply allometric scaling rules. Allometric scaling relates physiological and kinetic parameters to body weight raised to a certain power (for scaling according to caloric demand, the allometric exponent is 0.75).

In this report the evaluation of species differences in two large datasets is presented: NTP studies and repeated dose studies from REACH registrations, as already described in the report on “Time extrapolation”. In agreement with the predictions of allometric scaling the evaluation of oral studies from both datasets show that the larger species appears to be more susceptible, if doses are expressed per kg body weight (geometric means of dose ratio rats/mice: NTP: 0.40; REACH: 0.66). Inhalation studies with rats and mice (for which the most datasets are available) show a similar susceptibility (geometric means of dose ratio rats/mice: NTP: 0.96; REACH: 1.09), when exposure concentrations are compared. A relevant variability around the mean values is observed for all datasets (geometric standard deviations (GSDs): NTP: 3.6-3.8; REACH: 3.0-3.9).

These results are compared with existing empirical evaluations of toxicity (and toxicokinetic data). Although the associated uncertainty does not allow to determine whether an allometric exponent of 0.7 or 0.75 is more adequate, the existing evaluations support the application of allometric scaling factors. Caloric demand (also called metabolic rate) scaling is recommended here, as it is supported by the empirical data as well as by mechanistic considerations. Caloric demand scaling also leads to the conclusion that no correction is required when deriving an OEL from an inhalation concentration as point of departure.

Based on the existing data a two-step approach is recommended:

- Correction (normalisation) of doses by allometric scaling factors derived from caloric demand scaling (for oral data, no correction required for inhalation concentrations).
- Consideration of remaining uncertainty due to substance-to-substance variability by a distribution derived from the empirical datasets.

Due to its higher quality the empirical dataset derived from NTP data is preferable over the REACH dataset. However, both datasets include additional uncertainty from using NOAEL ratios (instead of BMD ratios). If possible, the uncertainty introduced by errors in using NOAELs for calculating ratios should be avoided.

Considering the high inherent uncertainty of the HEC procedure we propose to use the same distribution also to account for uncertainties in interspecies extrapolation in case of particulates assessed by the HEC procedure.

Abbreviations

AUC	Area under the curve
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
BMD(C)	Benchmark dose (concentration)
BW	Body weight
CI	Confidence Interval
Cmax	Maximum plasma concentration
ECHA	European Chemicals Agency
DNEL	Derived no effect level
EF/AF	Extrapolation factor /Assessment factor
EPA	Environmental Protection Agency (in the US)
FDA	Food and drug administration (in the US)
GM	Geometric mean
GSD	Geometric standard deviation
HEC	Human equivalent concentration
LD_{10}/LD_{50}	Dose corresponding to 10/50 % lethality
NO(A)EL/ NO(A)EC	No (adverse) effect level/concentration
NTP	National Toxicology Program

OEL	Occupational exposure limit
PBPK	Physiology-based pharmacokinetic (model)
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals, Regulation (EC) No 1907/2006
RTECS	Registry of Toxic Effects of Chemical Substances
US EPA	US Environmental Protection Agency
US FDA	US Food and Drug Administration
WHO	World Health Organization

1 Introduction

Animal models are used in toxicology, because humans share principal characteristics of anatomy and physiology with other mammals (Boxenbaum 1984). But already in the 1960ies scientists noted that there are regular relationships between physiological parameters such as clearance rates and oxygen consumption or other parameters indicating metabolic rate of species differing in size (Boxenbaum 1982). From that, allometric principles were described, which relate physiological and kinetic parameters to body weight raised to a certain power. For example, caloric demand (or metabolic rate) scaling predicts that physiological parameters such as clearance can be scaled between species of different sizes with body weight$^{0.75}$. Subsequently, allometric principles were increasingly used in regulatory toxicology, but also in pharmacology (Mahmood 2012; Sharma and McNeill 2009). These allometric theories will be explained in more detail in chapter 4.3.

But even after normalisation of doses expressed per kg body weight for differences in body size by allometric scaling animal models might show differences in susceptibility compared to humans (Sharma and McNeill 2009). Differences, which lead to animals showing higher or lower toxicity than humans at the same external dose, might be caused by quantitative differences in toxicokinetics. For example, variations in xenobiotic metabolising enzymes may lead to faster or slower metabolism of a xenobiotic substance and/or different metabolites (Griem et al. 2002; Mumtaz and Pohl 2012). Differences in response to toxic agents at the same internal dose can be caused by variation in repair mechanisms or differences in molecular receptors (toxicodynamic differences). Also, toxic effects are most often not caused by single events, but by cascades of processes including intercellular and intracellular signal transduction and at each step differences between species might occur, which result in different susceptibilities (Mumtaz and Pohl 2012).

In the following chapters we use the database created from NTP studies and repeated dose studies from REACH registration dossiers to analyse species differences. Methodological details on creation and evaluation of these databases were explained in the report on “Time extrapolation”. Therefore, we refer to this report for a description of the methods used. In chapter 3 the results for the evaluation of interspecies differences observed in these databases are presented, which are further discussed and compared in chapters 4.1 and 4.2. In the following parts of chapter 4 the results are compared with literature data and discussed with regard to their relevance for regulatory purposes.

2 Methods

The procedure to obtain ratio distributions which serve as the empirical basis to evaluate interspecies extrapolation factors is virtually identical to the analysis of the distributions for time extrapolation. All methodological steps for deriving the ratio distributions in this report are included in the report on time extrapolation. This also applies to explanations regarding the statistical methods and graphical representations. In short, study pairs are formed, where each pair consists of two studies with the same substance and the same length, but with different species. The ratios of the dose descriptors (either NOAELs or LOAELs) of the study pairs are calculated. Each ratio is derived in the way that the dose of the study with the species that has the higher body weight is the numerator.

In chapter 4.3 a theoretical introduction into the application of allometric rules in toxicological hazard assessment is given. In order to allow for easier interpretation of the results reported in the following chapter we provide here the expected values for

- Body weight scaling (no influence of size of species, doses expressed per kg body weight are expected to be equipotent, i.e. “scaling” according to body weight^{1})
- Caloric demand (or metabolic rate) scaling, as the most frequently used method (with scaling according to body weight$^{0.75}$).

Table 2-1 Expected ratios of equipotent doses per kg body weight for the comparison of species with different body weights, separated by exposure route

Species comparison	**Body weight scaling**		**Caloric demand scaling**	
	Oral	**Inhalation**	**Oral***	**Inhalation**
Rat/mouse	1	1	0.59	1
Dog/rat	1	1	0.34	1
Dog/mouse	1	1	0.20	1
Primate/rat	1	1	0.50	1

* body weight species 2$^{0.25}$/body weight species 1$^{0.25}$, calculated with body weights as given in Table 4-3 (see chapter 4.3)

3 Results

3.1 Evaluation NTP data

3.1.1 Influence of exposure route

Exposure route is an obvious factor to analyse interspecies differences as different ratio distributions are to be expected between oral and inhalation exposures due to allometric differences between species, which are assumed to be cancelled out in inhalation experiments by different breathing volumes. Indeed, the ratio of effects between rats and mice for inhalation studies is centred around 1[1]. The ratios for oral exposure are shifted in a way that the species with higher body mass is the more sensitive one[1] (meaning the dose descriptor for that species is lower) (Figure 3-1 and Table 3-1).

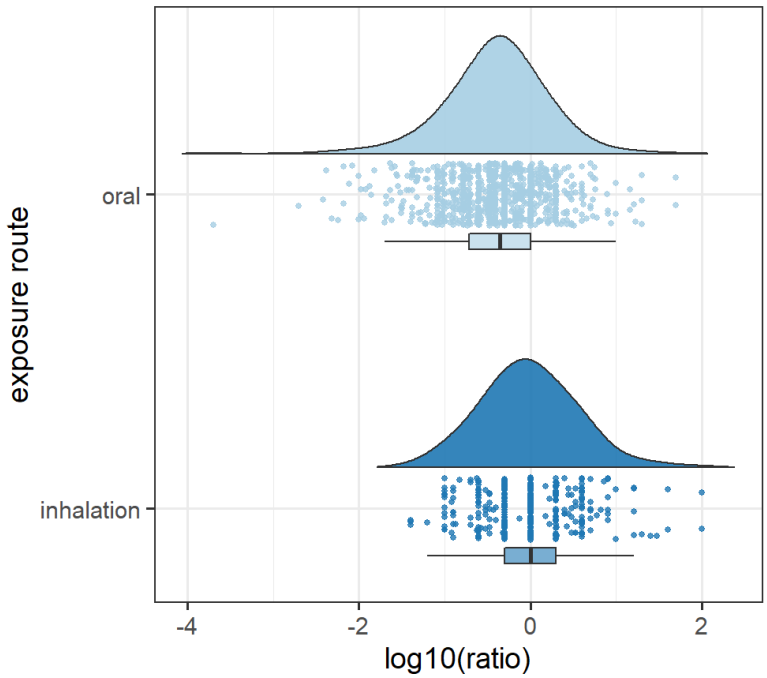


Figure 3-1 Distribution of ratios for the comparison of rats with mice, separated by exposure route

Because of this strong impact of exposure route on the ratio distribution, the influence of the other experimental factors on interspecies ratio distributions was always investigated on the dataset split into oral and inhalation exposure.

[1] 95% CI of the GM for oral exposure: 0.37 - 0.44, for inhalation exposure: 0.84 - 1.11

Table 3-1 Summary statistics of distributions of ratios for the comparison of rats with mice, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	0.99	2.92	0.40	3.78	0.04	0.44	1.00	2.97	927
inhalation	2.73	8.74	0.96	3.61	0.12	1.00	2.00	8.00	333

3.1.2 Influence of sex (by route)

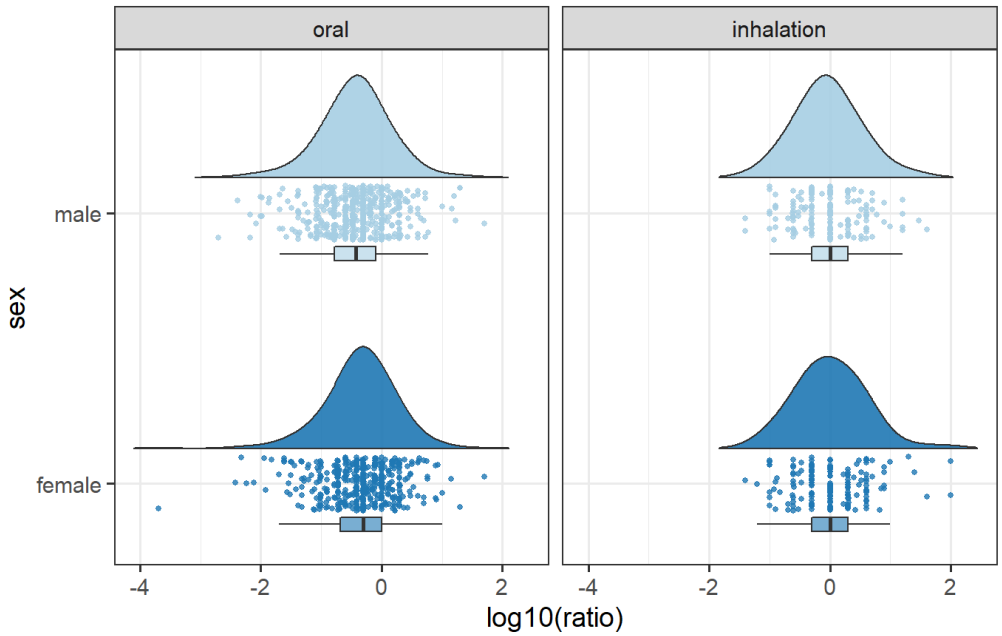


Figure 3-2 Distribution of ratios for the comparison of rats with mice, separated by route and sex

If the dataset split by route is additionally split by the sex, it is evident that there is no difference between males and females as the respective distributions are largely congruent (Figure 3-2 and Table 3-2). For oral exposures, GM and 75th percentile of the rat/mouse comparison for males seem slightly smaller than for females, however the 95% CI for these parameters are largely overlapping[2,3]. In consequence, there is

[2] 95% CI of the GM for males, oral exposure: 0.33 – 0.41, for females, oral exposure 0.39 – 0.50.
[3] 95% CI of the 75th percentile for males, oral exposure: 0.67 – 1.00, for females, oral exposure: 0.88 – 1.04

no reason for differentiating between sex for the interspecies ratio distribution between rats and mice.

Table 3-2 Summary statistics of distributions of ratios for the comparison of rats with mice, separated by route and sex

Route	Sex	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	**male**	0.96	2.97	0.37	3.75	0.04	0.37	0.80	2.67	484
oral	**female**	1.03	2.85	0.44	3.80	0.05	0.50	1.00	3.00	443
inhalation	**male**	2.15	4.57	0.91	3.44	0.12	1.00	2.00	7.30	168
inhalation	**female**	3.31	11.53	1.02	3.80	0.10	1.00	2.00	8.00	165

3.1.3 Influence of study duration (by route)

The study duration has no influence on the ratio distribution of the interspecies comparison between rats and mice. The distribution for subchronic studies is broader than for the other study durations, but this is strongly influenced by few extreme values. Conversely, ratios from studies with inhalation exposure of subacute duration have a more narrow distribution, but this could also be by chance because fewer extreme values occurred within the lower number of samples (Figure 3-3 and Tables 3-3).

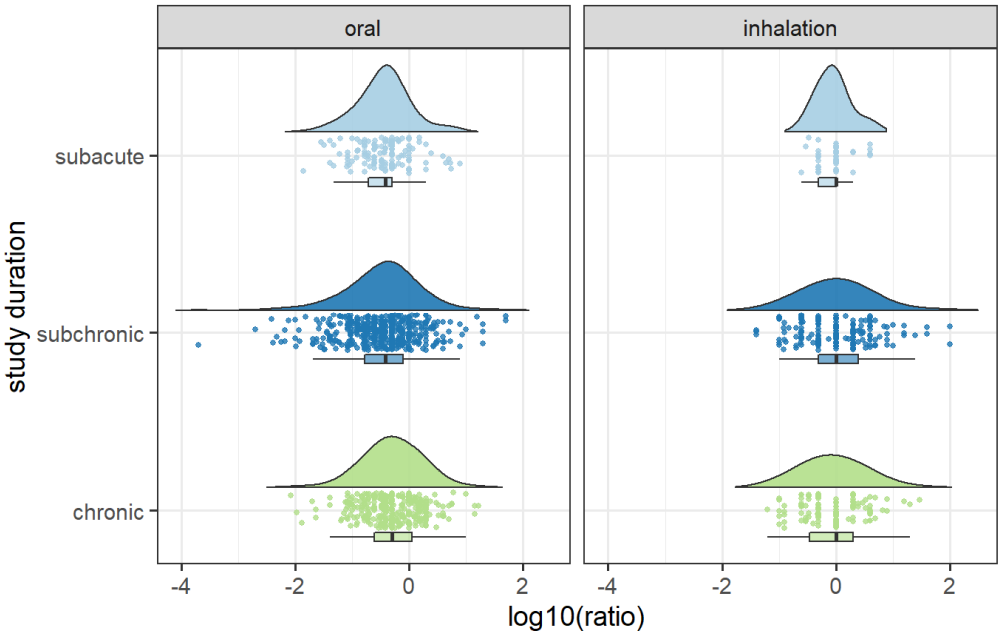


Figure 3-3 Distribution of ratios for the comparison of rats with mice, separated by study route and duration

The GM of ratios from inhalation subacute, subchronic and chronic studies are nearly identical. However, as the distribution of the subacute studies has a narrower shape the 75th percentile for this distribution is different from the distributions for subchronic or chronic studies[4]. For subacute studies only effects on bodyweight are compared, which might explain the difference to subchronic and chronic studies. The effect of endpoints on the distribution is evaluated in section 3.1.4.

For oral exposure, the GM and location of percentiles for ratios from chronic exposures seems to be slightly higher than the respective distributions for subacute and subchronic studies. This difference is significant according to the 95% CI of the GM[5] and 75th percentile[6], but the difference likely is too small to be of practical relevance.

[4] 95% CI of the 75th percentile for inhalation, subacute: 1.00 – 2.00, for inhalation, subchronic: 2.00 – 3.33, for inhalation, chronic: 1.67 – 3.33.
[5] 95% CI of GM for ratios of oral studies with subacute, subchronic and chronic exposure: 0.31 - 0.44, 0.30 – 0.40, 0.46 – 0.59.
[6] 95% CI of the 75th percentile for oral studies with subacute, subchronic and chronic exposure: 0.50 – 0.80, 0.69 – 1.00, chronic: 1.00 – 1.41.

Table 3-3 Summary statistics of distributions of ratios for the comparison of rats with mice, separated by route and study duration.

Route	Study duration	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	**subacute**	0.68	1.11	0.37	2.93	0.06	0.38	0.51	2.02	140
oral	**subchronic**	1.08	3.80	0.35	4.37	0.03	0.38	0.80	2.96	465
oral	**chronic**	1.01	1.75	0.51	3.21	0.08	0.50	1.14	3.00	322
inhalation	**subacute**	1.17	1.02	0.90	1.98	0.34	1.00	1.00	4.00	42
inhalation	**subchronic**	3.58	11.54	1.03	4.14	0.10	1.00	2.50	10.00	177
inhalation	**chronic**	1.98	3.79	0.89	3.43	0.12	1.00	2.00	5.58	114

3.1.4 Influence of toxicity endpoint (by route)

For the oral path, no local effects have been evaluated for the NTP data. The available distributions from oral studies for the endpoints body weight and systemic effects show only insignificant differences. In the case of inhalation studies, the GM for ratios derived from effects on bodyweight is slightly shifted towards greater ratio values (1.26, i.e. a higher sensitivity of mice), whereas for local and systemic effects the GM was < 1.0, indicating that rats were more sensitive.[7]. Because all three CI for inhalation studies comprise the ratio of 1 or are very close to 1 (on the normal scale) and the differences in the distribution are rather small, this is likely not of further relevance. The median for inhalation studies of all three endpoints is 1.

[7] 95% CI of the GM for inhalation, bodyweight: 1.03 – 1.54, for inhalation, local: 0.61 – 1.01, for inhalation, systemic: 0.58 – 1.00.

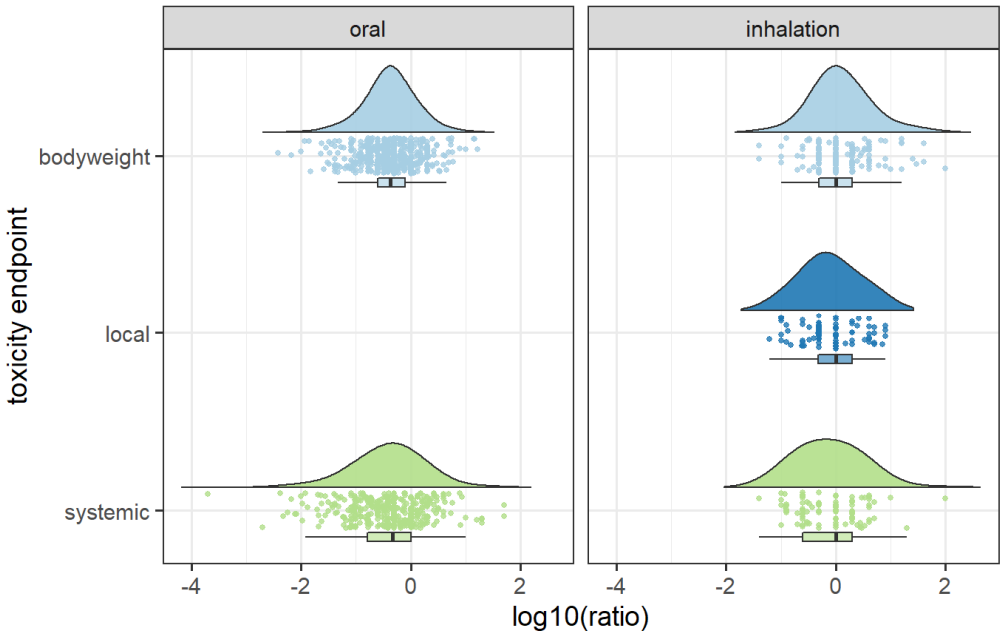


Figure 3-4 Distribution of ratios for the comparison of rats with mice, separated by study route and endpoint

Table 3-4 Summary statistics of distributions of ratios for the comparison of rats with mice, separated by route and endpoint

Route	Endpoint	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	**bodyweight**	0.80	1.46	0.42	3.12	0.06	0.42	0.80	2.08	549
oral	**systemic**	1.28	4.20	0.38	4.80	0.02	0.47	1.00	3.99	378
inhalation	**bodyweight**	3.50	9.86	1.26	3.51	0.23	1.00	2.02	16.00	152
inhalation	**local**	1.54	1.93	0.79	3.29	0.10	1.00	2.00	5.00	85
inhalation	**systemic**	2.55	10.35	0.76	3.82	0.12	1.00	2.00	4.25	96

3.1.5 Stratification according to target organs (by route)

It was investigated whether substances that evoke liver or kidney toxicity lead to different ratio distributions than substances with other target organs. All analyses were

performed on the dataset split by route. Ratios from substances which had both kidney and liver as target organs (applies only to 3 ratios) are a part of both the “kidney” and the “liver” group.

Table 3-5 Summary statistics of distributions of ratios for the comparison of rats with mice, separated by route and target organ

rRute	Target organ	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	**other**	1.00	2.99	0.40	3.66	0.04	0.40	0.90	3.00	831
oral	**liver**	1.05	2.53	0.41	5.30	0.04	0.50	1.02	2.00	64
oral	**kidney**	0.83	1.08	0.40	4.25	0.05	0.50	1.00	2.82	32
inhalation	**other**	2.78	8.93	0.97	3.61	0.12	1.00	2.00	8.00	318
inhalation	**liver**	1.48	3.02	0.52	4.15	0.13	0.50	1.00	5.95	10
inhalation	**kidney**	2.20	1.75	1.66	2.42	0.60	2.00	2.50	4.50	5

For both exposure routes, no differences in the distributions can be found (**Table 3-5**). For the oral route, this is immediately evident, as the percentiles are very close together. For inhalation studies, the number of ratios from substances with kidney or liver effects is too small to make any meaningful conclusions.

It was further checked whether any differences appear in the distributions for oral studies when the data is additionally separated by exposure duration. According to the data, the study duration has no significant influence on ratio distributions (no data shown).

3.1.6 Stratification according to chemical features (by route)

The influence of certain classes of chemicals on the distributions was investigated with the same groups of chemicals as for the time extrapolations (see table 3-12 in the report on “Time extrapolation”). According to the impact of exposure route on the distribution parameters, the analyses was performed on the dataset split into ratio distributions derived from oral and from inhalation studies (Table 3-6).

Table 3-6 Summary statistics of distributions of ratios for the comparison of rats with mice, separated by route and substance groups

Route	Substance group	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	**other**	1.00	2.96	0.40	3.81	0.04	0.45	1.00	3.00	900
oral	**alkylated aromatics**	1.23	1.55	0.49	5.16	0.09	0.58	1.75	3.50	6
oral	**metal compounds**	0.58	0.51	0.41	2.39	0.11	0.33	0.77	1.50	21
inhalation	**other**	2.15	6.75	1.00	3.07	0.20	1.00	2.00	5.00	245
inhalation	**alkylated aromatics**	5.98	10.55	1.57[8]	5.76	0.10	1.00	5.00[9]	32.00	37
inhalation	**metal compounds**	3.14	13.98	0.56[8]	4.33	0.10	0.50	1.00[9]	7.00	51

For oral exposure, GM values don't show an influence of chemical classes. In addition, the number of ratios derived from studies with alkylated aromatics or metal compounds is rather low.

For inhalation exposure the number of ratios ("n") for the alkylated and metal compounds is higher and indeed there seem to be substance group-specific differences in sensitivity. For alkylated aromatics, the parameters of the distribution are shifted towards a higher sensitivity of mice (GM: 1.57, 75th percentile: 5.00), yet the statistical evaluation indicates that this shift is only significant at the 95% confidence level regarding the 75th percentile[9], but not the GM[8]. For metal compounds it is the other way round: rats are more sensitive than mice, and only regarding the GM (GM: 0.56[8], 75th percentile: 1.00[9]).

To follow up on these differences after inhalation exposure, we additionally looked at the influence of the other experimental factors. The sex of the test animals, although reported to be an influencing factor for species sensitivity for specific substances Kratchman et al. (2018) did not have such an effect according to our analysis (no data

[8] 95% CI of the GM for inhalation, alkylated aromatics: 0.90 – 2.76, for inhalation, metal compounds: 0.38 – 0.85.

[9] 95% CI of the 75th percentile for inhalation, alkylated aromatics: 3.33 – 8.00, for inhalation, metal compounds: 1.00 – 3.00.

shown). Neither did the separate analysis by study duration reveal any new distinctive features (data not shown).

Splitting the data by type of endpoint (bodyweight, systemic, local) pronounces the species differences regarding the substance classes. As reported above, mice showed a higher sensitivity towards inhalation exposure with the alkylated aromatics (Table 3-6). This seems to be primarily hailing from effects on bodyweight, as these sensitivity differences don't show up for comparisons of systemic toxicity and local toxicity in the lung (Table 3-7) and correspondingly the GM for the effects on bodyweight is shifted especially far away from 1 (4.44). While it is well possible that a substance exacerbates species specific effects only on specific endpoints, the low sample size has to be taken into consideration. Yet, the 95% CI, obtained by bootstrapping the empirical distribution, indeed indicates that the distribution is shifted towards values greater than 1[10]. In consequence there is quite a high likelihood that the sensitivity regarding bodyweight is indeed shifted towards mice. The situation for the higher sensitivity of rats towards metal compounds is very much the same. This effect could already be seen with all endpoints pooled but is stronger when looking at local effects only[11]. Again, the low number of ratios need to be taken into account.

[10] 95% CI of the GM for exposure to alkylated aromatics, effects on bodyweight: 2.25 – 8.78.
[11] 95% CI of the GM for exposure to metal compounds, local effects: 0.22 – 0.64.

Table 3-7 Summary statistics of distributions of ratios for the comparison of rats with mice, separated by endpoint as well as substance groups (inhalation only)

Substance group	Endpoint	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
other	**bodyweight**	2.78	9.67	1.16	2.99	0.25	1.00	2.00	6.90	112
other	**local**	1.57	1.77	0.87	3.09	0.13	1.00	2.00	5.00	69
other	**systemic**	1.68	2.65	0.90	3.13	0.20	1.00	2.00	4.00	64
alkylated aromatics	**bodyweight**	10.96	13.94	4.44	4.41	0.87	4.00	16.00	40.00	17
alkylated aromatics	**local**	2.92	3.95	0.77	7.41	0.10	0.67	6.25	8.00	6
alkylated aromatics	**systemic**	1.24	1.56	0.60	3.55	0.10	0.44	1.75	3.92	14
metal compounds	**bodyweight**	1.52	2.16	0.75	3.52	0.10	1.00	1.33	3.93	23
metal compounds	**local**	0.52	0.37	0.38[11]	2.42	0.10	0.49	0.88	1.00	10
metal compounds	**systemic**	6.67	23.41	0.49	6.73	0.10	0.25	1.00	23.50	18

3.2 Evaluation REACH data

In contrast to the NTP data, which is restricted to the two species rats and mice, the REACH dataset also contains study results from other species (e.g. dogs and primates). This is of potentially high value for scrutinizing the applicability of allometric principles, which assume a universal scaling rule across all species. Data on species comparisons other than rats vs mice might be used for additional evaluation regarding applicability of allometry as it is used for regulatory purposes. However, not surprisingly, most studies reported under REACH are conducted on rats and mice. Studies on other species are also predominantly older studies, which more often do

not meet the selection criteria, thus leading to a very low number of valid pairs suitable for deriving a ratio. Following rats and mice, the next most common species combination is rats and dogs, which resulted in 72 ratios for oral exposure and 7 ratios for inhalation exposure in our evaluation. This species combination is included in the evaluations following below, although the low numbers often prevent statistically meaningful conclusions. Additionally, rare species combinations were possible based on the processed data (dog/mouse: 25 ratios, rat/primate: 13 ratios, rat/pig: 1 ratio). Although the low number of ratios is limiting possible interpretations, the dog/mouse and primate/rat comparisons are based on species with a high body weight difference and are of particular value and are therefore reported below.

3.2.1 Influence of exposure route

3.2.1.1 Rat vs mouse

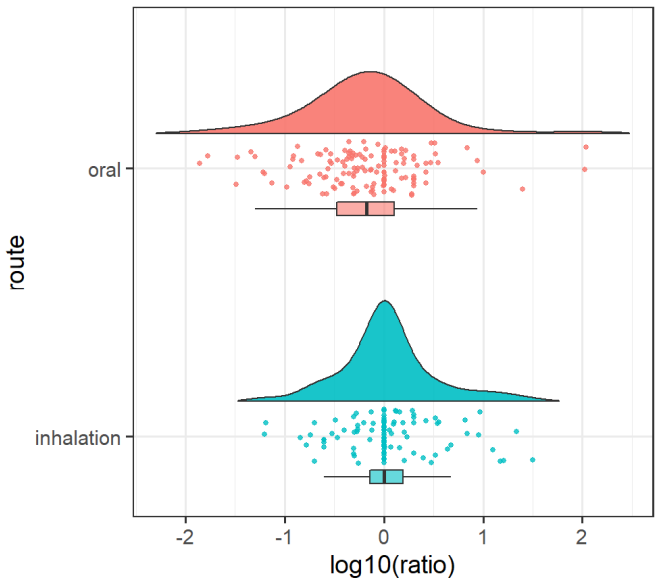


Figure 3-5 Distribution of ratios for the comparison of rats and mice, separated by exposure route

As already observed for the NTP data, following the allometric principle, differences between exposure routes are also observed for the REACH data (Figure 3-5). In case of oral data the GM is shifted towards rats (the species with higher body mass) being the more sensitive species[12]. The distribution for inhalation exposure is centred around 1, as would have been expected[12] (Table 3-8). A slightly higher GSD for oral data compared to inhalation data was observed.

[12] 95% CI of the GM for oral exposure: 0.52 – 0.83, for inhalation exposure: 0.88-1.34

Because of this impact of the route on the ratios, the analysis of the other experimental factor in the REACH dataset, the study duration, is split by route.

Table 3-8 Summary statistics for distributions of ratios for the comparison of rats with mice, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	2.82	13.21	0.66	3.85	0.06	0.67	1.27	3.39	135
inhalation	2.28	4.44	1.09	2.98	0.20	1.00	1.56	9.28	105

3.2.1.2 Dog vs rat

For the comparison of dogs vs rats (Figure 3-6 and Table 3-9) only very few datapoints for the inhalation route are available (7 ratios) which renders a comparison of routes not meaningful. The oral data however shows the characteristics that would be expected for two species with different bodyweights. The ratio distribution is shifted towards dogs as the more sensitive species, i.e. towards ratios < 1 with a GM of 0.68[13]. The value is higher than the value expected according to caloric demand scaling (0.34, see chapter 2). However, it should be noted that dog studies in the REACH database are often chronic studies with a duration of 1 - 2 years. For calculation of the ratios, these are compared to chronic studies in the rat (typically 2 years), which represent a much larger fraction of the animal lifespan.

[13] 95% CI of the GM for oral exposure: 0.48 – 0.98

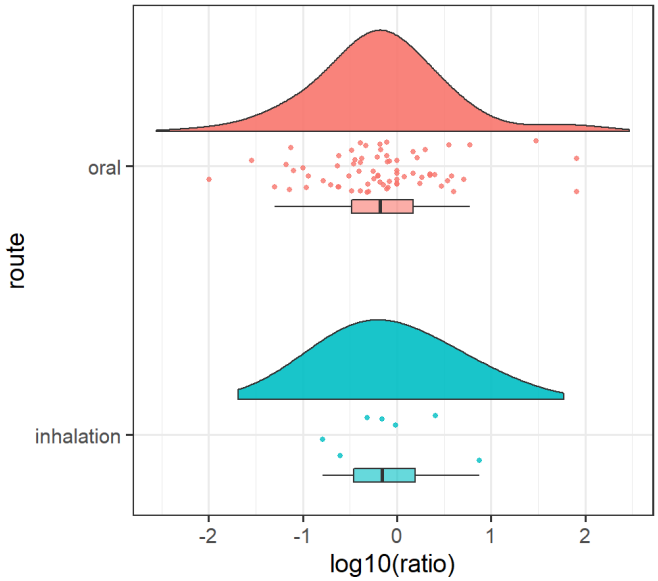


Figure 3-6 Distribution of ratios for the comparison of dogs with rats, separated by exposure route

Table 3-9 Summary statistics for distributions of ratios for the comparison of dogs with rats, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	3.69	13.73	0.68	4.71	0.07	0.66	1.50	5.54	72
inhalation	1.80	2.64	0.82	3.77	0.19	0.70	1.76	6.02	7

3.2.1.3 Dog vs mouse

Only 16 ratios for oral and four ratios for inhalation exposures could be derived for the species comparison dog vs mouse (Figure 3-7 and Table 3-10). The different lifespan coverage of dog and mouse studies, as explained in the dog vs rat section above, also applies here, yet the GM for oral exposure (0.26) fits quite well to the expected value of 0.20. However, in light of the low amount of datapoints this GM has a relatively high uncertainty, as is reflected by the wide 95% CI (0.10 – 0.63). The GM or the inhalation exposure does not contradict the expected value of 1.00, but based on only four values no meaningful conclusions can be drawn.

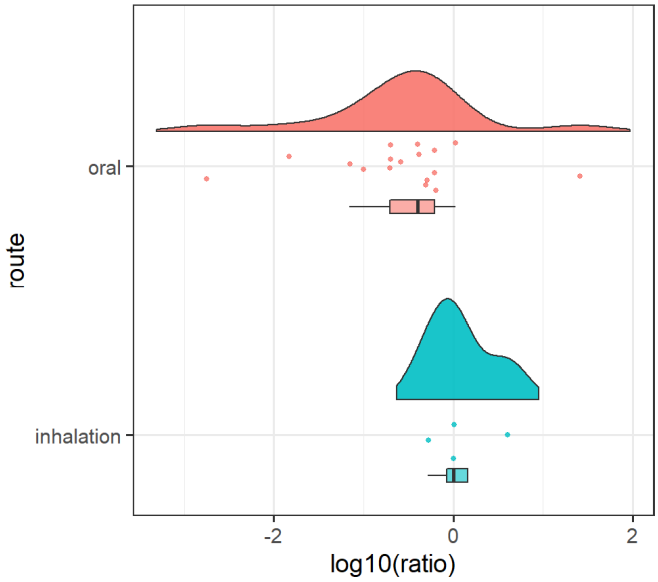


Figure 3-7 Distribution of ratios for the comparison of dogs with mice, separated by exposure route

Table 3-10 Summary statistics for distributions of ratios for the comparison of dogs with mice, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	1.86	6.17	0.26	7.23	0.01	0.40	0.61	6.00	17
inhalation	1.64	1.60	1.21	2.35	0.60	1.01	1.77	3.56	4

3.2.1.4 Primate vs rat

For the comparison primate vs rat only 6 ratios from oral and 4 ratios from inhalation studies could be derived with the selected quality criteria (Figure 3-8 and Table 3-11). The GM of oral exposures (3.14) and of inhalation studies (2.63) is quite off from the expected values (0.50 and 1.00, respectively). Because these parameters are based on such a low number of ratios, no meaningful interpretation is possible.

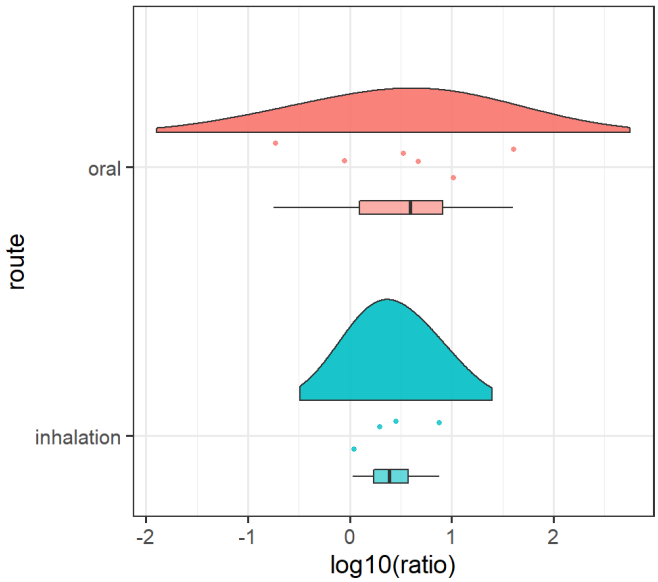


Figure 3-8 Distribution of ratios for the comparison of primates with rats, separated by exposure route

Table 3-11 Summary statistics for distributions of ratios for the comparison of primates with rats, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	9.81	15.15	3.14	6.57	0.36	3.93	8.63	32.42	6
inhalation	3.39	2.85	2.63	2.26	1.21	2.48	4.10	6.82	4

3.2.2 Influence of study duration route (by route)

To analyse whether, in addition to the route, the study duration has an influence on the ratio distributions, the ratios of the rat vs mouse comparison were additionally split by the study duration. This analysis was not performed for other datasets, as for the low number of ratios no meaningful result was expected.

It appears that after oral exposure, there is no difference in the sensitivity of rats and mice in subacute studies (GM 1.03). However, this distribution has a large spread and the statistical analysis indeed indicates that the data does not permit such a conclusion. The CI of the GM are overlapping for all 3 study durations, indicating no differences for

oral exposure[14]. Further, the 75th percentiles of the distributions after oral exposure are not different from each other[15]. For the inhalation route, the ratios are centred around 1 and no differences are observed between the study durations[16,17](Figure 3-9 and Table 3-10).

In conclusion, the data indicate that there is no need to split the ratio distributions by study duration for the further analyses.

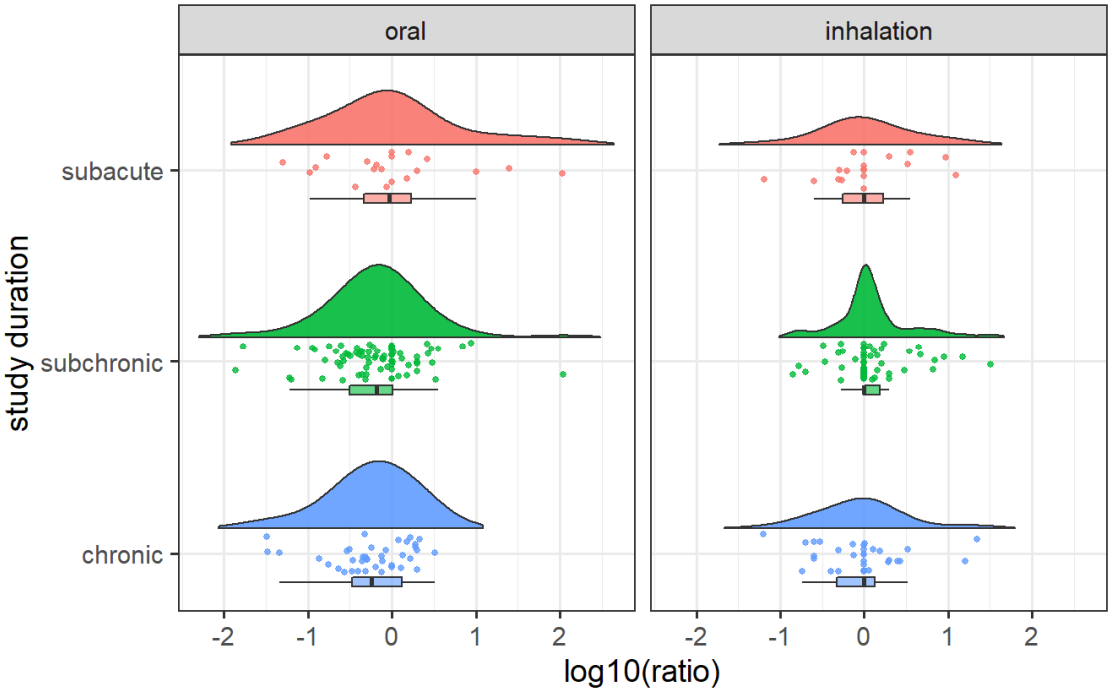


Figure 3-9 Distribution of ratios for the comparison of rats and mice, separated by exposure route and study duration

[14] 95% CI of the GM for oral subacute studies: 0.48-2.30, for oral subchronic studies: 0.47-0.85, for oral chronic studies: 0.40 – 0.78

[15] 95% CI of the 75th percentile for oral subacute studies: 1.00 – 13.77, for oral subchronic studies: 1.00 – 1.99, for oral chronic studies: 0.77 – 1.91

[16] 95% CI of the GM for inhalation subacute studies: 0.61-1.85, subchronic: 0.95-1.60, chronic: 0.61-1.37

[17] 95% CI of the 75th percentile for inhalation subacute studies: 1.00 – 13.77, for inhalation subchronic studies: 1.00 – 1.99, for inhalation chronic studies: 0.77 – 1.91

Table 3-12 Summary statistics for distributions of ratios for the comparison of rats and mice, separated by exposure route and study duration

Route	Study duration	Mean	SD	GM	GSD	5%	Median (50%)	75%	95%	n
oral	**subacute**	7.83	23.91	1.03	6.16	0.10	0.93	1.70	29.16	20
oral	**subchronic**	2.50	12.57	0.63[14]	3.73	0.07	0.65	1.03	3.38	76
oral	**chronic**	0.88	0.74	0.56	3.02	0.04	0.57	1.31	2.02	39
inhalation	**subacute**	2.22	3.35	1.06	3.44	0.22	1.00	1.75	9.82	18
inhalation	**subchronic**	2.38	4.75	1.23	2.69	0.29	1.00	1.52[17]	7.58	55
inhalation	**chronic**	2.15	4.54	0.90	3.28	0.19	1.00	1.34	9.04	32

4 Discussion

4.1 Methodological approach

Most of the methodological discussion of the distributions for time extrapolation also applies to the interspecies comparison as the data were derived from the same data sources with the same selection criteria and same processing steps. Yet, the comparison is always performed on studies with the same duration. This has the consequence that some characteristics of the data have a lower potential to impact results than they have for time extrapolation. Such characteristics are the number of animals and the dose spacing. As these parameters depend primarily on the study duration they are very comparable, if not identical, for the vast majority of compared study pairs and the possibility of an influence on the resulting ratio can be considered low.

4.2 NTP vs REACH

An overview of the distributions derived from NTP and REACH data is given in Table 4-1. The reasoning regarding data quality that leads to preference of the distributions derived from NTP data over the distributions derived from REACH data for evaluating the time extrapolation basically also applies to the evaluation of factors for interspecies extrapolation in this report.

In addition, because only one tested species is needed for REACH registrations, the amount of valid ratio pairs that can be used for comparison is drastically reduced in comparison to NTP studies, where nearly all substances were tested on rats and mice. The low number of ratios from the REACH data is particularly striking for subacute studies, where less ratios could be calculated than for subchronic and chronic studies. The REACH data potentially provide additional value by potentially allowing comparisons between other species than just rats and mice. Unfortunately, the amount of valid ratio pairs among other species is too low to provide meaningful insight.

Table 4-1 Overview of interspecies comparison based on NTP and REACH data

Species	Route	Study source	N study pairs	GM (95% CI)	Expected value *	GSD	75% percentile (95% CI)
rat/ mouse	oral	NTP	927	0.40 (0.37-0.44)	0.59	3.78	1.00 (0.78-1.00)
rat/ mouse	oral	REACH	135	0.66 (0.52-0.83)	0.59	3.85	1.27 (1.00-1.67)
rat/ mouse	Inha-lation	NTP	333	0.96 (0.84-1.10)	1.00	3.61	2.00 (2.00-2.50)
rat/ mouse	Inha-lation	REACH	105	1.09 (0.88-1.34)	1.00	2.98	1.56 (1.14-2.00)
dog/rat	oral	REACH	72	0.68 (0.48-0.98)	0.34	4.71	1.50 (0.88-2.26)
dog/rat	Inha-lation	REACH	7**	(0.82)	1.00	3.77	(1.76)
dog/ mouse	oral	REACH	17	0.26 (0.10-0.63)	0.20	7.23	0.61 (0.40-1.06)
dog/ mouse	Inha-lation	REACH	4**	(1.21)	1.00	2.35	(1.77)
primate/ rat	oral	REACH	6**	(3.14)	0.50	6.57	(8.63)
primate/ rat	Inha-lation	REACH	4**	(2.63)	1.00	2.26	(4.10)

*According to caloric demand scaling, calculated with body weights as given in Table 4-3; see chapter 4.3 for algorithm

** Not enough ratios to be meaningful

4.3 The principles of allometric scaling

It is for many years now that scientists observed that certain physiological parameters vary between species in a regular way dependent on body size or body weight (Boxenbaum 1982). For example, heart rates or respiratory rates are much faster in smaller species compared to man. These observations were described by allometric equations, which relate such parameters to body weight raised to an exponent b:

$$Y = a\,W^{b}$$

Or

$$logY = \log a + (b\ \log W)$$

where Y is the (physiological) parameter under consideration, a is proportionality constant and W is the body weight.

It was observed that volumes and weights show relationships between species with an exponent b=1 (such as body weight itself), whereas rates (masses or volumes moved per time unit) follow an exponent around ¾ (Kenyon 2012). Exponents around -1/4 result, when parameters which scale with exponent ¾ are normalised to body weight ($BW^{3/4}/BW=BW^{-1/4}$) (US EPA 2011). Table 4-2 shows some of the relationships observed.

Table 4-2 Body weight scaling rules for physiological parameters (adapted from (US EPA 2011)

Parameter	**Unit**	**Scaling rule**
Energy utilization	J/d	$BW^{3/4}$
Glomerular filtration rate	L/min	$BW^{3/4}$
Glucose turnover	mg/min	$BW^{3/4}$
Food consumption	g/d	$BW^{3/4}$
Water consumption	L/d	$BW^{3/4}$
Heart rate	1/min	$BW^{-1/4}$
Respiratory rate	1/min	$BW^{-0.26}$
Blood volume	L	BW^{1}
Vital lung capacity	mL	BW^{1}

From these observations the concept of physiological time (also called caloric demand or metabolic rate scaling) was derived: The relationship of physiological parameters between species of differing size is not controlled by body weight, but by the speed of

physiological processes (“physiological time”), the key indicator of which is the energy consumption of the organism (“caloric demand scaling” or “metabolic rate scaling”[18]).

US EPA (2011) is giving an example to illustrate these relationships: There is an approx. 2300-fold difference in body mass and absolute heart mass between humans and mice (scaling to BW^{1}). However, cardiac output in humans is only about 300-fold greater than in mice (scaling to $BW^{3/4}$), whereas the heart rate in humans is about 7-fold less than in mice (scaling to $BW^{-1/4}$).

In the following, this concept of caloric demand scaling was expanded to toxicological questions, e.g. by Boxenbaum (Boxenbaum 1982, 1984), who investigated how clearance data for pharmaceuticals from several species comply with allometric relationships. Travis and White (1988) were among the first to apply the concept to toxicity. They compared data on the toxicity of antineoplastic agents in various species. Despite some uncertainties these empirical evaluations confirmed the applicability of caloric demand scaling in toxicology. An overview on existing empirical evaluations and a comparison with the data evaluation in this project is given in chapter 4.5, below.

As food and water consumption scale with exponent ¾, no scaling would be required, if comparison between species is done based on concentrations in food/feed or drinking water. But when doses are expressed per kg body weight, correction is required to conclude on equipotent human doses, as chronic exposure to 1 mg/kg bw/day in the mouse is not expected to cause the same effects as in humans.

For the same reason **no** allometric correction is required, if a health-based standard is derived from air concentrations in an inhalation study. As ventilation rates scale with an exponent of ¾ (Kenyon 2012), at a given air concentration smaller species take up more substance per kg body weight (and eliminates faster) than humans, leading to similar internal exposures.

It is important to note that allometric scaling is a normalisation procedure, which allows us to continue expressing oral doses as amounts per kg body weight. From the allometric equations species-specific so-called scaling factors can be derived, which can be used to calculate equipotent human doses from experimental animal data (see Table 4-3 for scaling factors currently used for regulatory purposes).

An oral dose expressed per kg body weight can be adjusted by scaling as follows:

Equation 1: $$\mathrm{dose}_{\mathrm{species\ 2}} = \frac{\mathrm{dose}_{\mathrm{species\ 1}}}{\mathrm{Scaling\ factor}}$$

Please note that although it is called a scaling “factor”, the dose of species 1 is actually divided by the scaling factor.

[18] Note that metabolic rate here means the basal metabolic rate (i.e. the rate of energy expenditure per unit time) of the organism and has nothing to do with metabolism of xenobiotics.

The scaling factor according to caloric demand scaling can be calculated as follows (see also ECHA (2012)):

Equation 2: $$\text{Scaling factor} = \frac{\frac{bw_{species\ 2}}{bw_{species\ 1}}}{\frac{bw_{species\ 2}^{0.75}}{bw_{species\ 1}^{0.75}}} = \frac{bw_{species\ 2}^{0.25}}{bw_{species\ 1}^{0.25}}$$

Examples (with body weights as given in Table 4-3):

Scaling from rat data (species 1) to humans (species 2):	4.1
Scaling from rat (species 1) to mouse (species 2):	0.59
Scaling from mouse (species 1) to rat (species 2):	1.7

Allometric scaling is not intending to account for variability between species due to differences in toxicokinetics. This substance-to-substance variability between species needs separate consideration (see chapter 4.5). As with all defaults, interspecies assessment factors (including scaling factors) can be replaced by substance-specific information, e.g. from PBPK models providing adequate modelling of species differences. Note that also for PBPK models allometric relationships are often used to scale parameters and to parameterize models if individual physiological values are unknown.

Several authors discussed the potential applicability domain of allometric scaling rules regarding the type of metabolic activation and type of toxic effect (acute versus chronic effects).

Travis et al. (1990) applied a simple PBPK model to investigate for various mechanistic cases, which allometric rules would be applicable:

- Direct acting compounds (not requiring metabolic activation)
- Reactive metabolite
- Stable metabolite (not undergoing metabolic transformation).

Based on these theoretical considerations the authors concluded that in all cases caloric demand scaling is adequate to predict toxicity if the toxic moiety is metabolically deactivated. Only for reactive metabolites which are spontaneously deactivated, body weight scaling seemed to be more adequate (see also Travis 1990). The underlying assumption of their modelling exercises was that the area under curve (AUC) is the most appropriate measure of internal dose and correlates best with toxicity. Several publications actually found a good agreement of data on clearance in various species with caloric demand scaling (Hu and Hayton 2001; Schneider et al. 2004), which confirms the assumption that also AUC ratios between species follows caloric demand scaling. No empirical data evaluation exists so far to support the exclusion of spontaneously deactivated metabolites. Even if the AUC of this metabolite follows other kinetics, the kinetics of the parent compound (uptake and metabolism) would most likely be rate-limiting and determine the delivery of the reactive metabolite to the target tissue.

In the report by US EPA (2011) it is discussed whether for short-term effects, peak exposures, which are better reflected by the maximum plasma concentration (Cmax), might be a better dose metric than AUC. However, no specific scaling rules for Cmax are suggested. Based on an evaluation of LD50 data from the database Registry of Toxic Effects of Chemical Substances (RTECS) Rhomberg and Wolff (1998) concluded that an exponent of 1 is more adequate for acute toxicity data. However, Schneider et al. (2004) demonstrated that this observation is an artefact resulting from RTECS reporting only the lowest LD50 per species. In fact, with a dataset from industrial chemicals the authors could show that acute toxicity data fitted better to caloric demand scaling than to body weight scaling with exponent 1. Nevertheless, this statement continues to be repeated in several review publications (Kenyon 2012; US EPA 2011). The same critique applies to the study of Burzala-Kowalczyk and Jongbloed (2011), who evaluated the same dataset.

Data from pharmacology, where relatively high numbers of data-rich active substances exist, might allow refining and improving allometric concepts, e.g. by establishing substance-specific correlations for clearance using data from various species or by considering plasma protein binding (Mahmood 2012; Poulin and Arnett 2017; Sharma and McNeill 2009). Other authors try to improve allometric relationships by adding additional influencing factors, such as body temperature, to the equations (Cao et al. 2014).

4.4 Comparison with regulatory practice

The US Environmental Protection Agency (US EPA) was the first to apply scaling rules to correct for the size of species. For the evaluation of carcinogens, prior to 1992 scaling according to body surface (body weight raised to the power 2/3) was used, but this practice was abandoned in favour of caloric demand scaling (US EPA 2011). Body surface scaling is still recommended by the US Food and Drug Administration (US FDA) for determining starting doses in clinical trials from toxicity observed in experimental animals, because scaling based on body surface results in a more conservative starting dose estimate (FDA 2005).

As discussed in the separate report “Comparison of methods for deriving OELs” in the EU all systems for deriving OELs or OEL-analogue values are recommending caloric demand scaling with an exponent of ¾, except the frameworks for assessing pesticides and biocides. These systems adhere to the assessment factor of 10, as also adopted by WHO (2009).

In the following table the scaling factors derived from caloric demand scaling and adopted for use for REACH purposes are shown.

Table 4-3 Scaling factors according to ECHA REACH Guidance for deriving DNELs for humans (with 70 kg body weight) from experimental studies with oral administration (adapted from ECHA 2012)

Species	**Body weight (kg)**	**Allometric scaling factor**
Rat	0.25	4
Mouse	0.03	7
Hamster	0.11	5
Guinea pig	0.8	3
Rabbit	2	2.4
Monkey	4	2
Dog	18	1.4

4.5 Comparison with published data

4.5.1 Toxicokinetic data

Schneider et al. (2004) identified sets of toxicokinetic data in literature for 71 substances comprising data for at least three different mammalian species. They analysed species differences for maximum plasma concentration, AUC, total clearance, elimination half-life and the volume of distribution. In regression analysis, they found a good agreement with predictions by caloric demand scaling. For example, for total clearance the median of the slopes of the substance-specific linear regressions was 0.76 (value predicted by caloric demand scaling: 0.75, equivalent with the allometric exponent), for AUC the observed median was 0.24 (compared to predicted value 0.25 (1-exponent 0.75)). For the volume of distribution the observed value was 0.96, compared to a predicted value of 1 (volumes and masses scale with an exponent of 1, i.e. they correlate linearly with bodyweight). Only for C_{max} the observed slopes (median 0.10) were more close to the prediction of body weight scaling (predicted value: 0, with exponent 1) than to the value predicted by caloric demand scaling (0.25). The latter might be taken as indication that short-term toxicity governed by peak exposures might not follow caloric demand scaling. However, the number of values for C_{max} in this study was low.

In a similar way Bokkers analysed toxicokinetic data (parameters AUC, clearance, elimination half-life and volume of distribution) (Bokkers 2009). Data for 159 to 319 (depending on the toxicokinetic parameter) for two to six species per compound were compiled from various unpublished and published sources, including Schneider et al. (2004). The author obtained arithmetic means (separate means were calculated for data coming from 2, 3, 4, etc. different species) for AUC of 0.32 – 0.34, for clearance of 0.71 – 0.75, for elimination half-life of 0.24 – 0.32, and the volumes of distribution of 0.96 – 1.03. These results are in agreement with an allometric scaling exponent in the range 0.7 to 0.75.

Huang et al. (2015) investigated interspecies differences regarding pharmacokinetics of 85 pharmaceutical substances. The investigated pharmacokinetic parameters were the time for systemic clearance and the volume of distribution at steady state. The value of this study for evaluating factors for interspecies extrapolation for toxicological assessment primarily is the large number of species for which pharmacokinetic data is available for the same substance. According to this analysis, the allometric exponent for systemic clearance ranges from 0.53 to 1.18, with a mean of 0.87, which is slightly higher than the value reported by Schneider et al. (2004). The allometric exponent for the parameter volume of distribution at steady status had a mean of 0.99, i.e. it scales linearly with bodyweight, as expected. The authors also compared their results with known pharmacokinetic values for humans and concluded that the exponent obtained in their analysis was too high for humans for most substances.

4.5.2 Toxicity data

4.5.2.1 Existing empirical evaluations

Apart from the set of toxicokinetic data described above Schneider et al. (2004) analysed three sets of toxicity data:

- Oral LD50 values from a set of toxicity data from industrial chemicals (8 mammalian species)
- NOAELs and LOAELs for rats, mice and dogs from long-term studies on pesticides
- Toxicity data (maximum tolerated dose and similar values) on anti-neoplastic agents after parenteral, 5-day application in six species including humans

When the acute toxicity data were compared, the data fit only moderately well to scaling according to the caloric demand. While for some species the median of the derived ratios corresponds quite well with caloric demand scaling, for the data-rich comparison mouse/rat it does not (ratio 0.86 instead of value 1.85 according to caloric demand scaling). For the comparison rabbit to rat a ratio of 0.55 was obtained, which is even lower than the expected ratio of 0.76. The authors could show that if always the lowest LD50 value per species was used for the calculations then the ratios tend to be closer to 1. From this observation they concluded that the statement made by Rhomberg and Wolff (1998) after their analysis of LD50 values from the RTECS database that an exponent of 1 is more adequate for acute toxicity data, is biased by

the data selection in RTECS (see chapter 4.3).The dataset on long-term effects of pesticides derived medians of the species ratios which fit clearly better to a scaling exponent of 0.75 than to an exponent of 1. Finally, the dataset on anti-neoplastic agents derived values which fit again very well to a scaling according to caloric demand. Taken together, the evaluations by Schneider et al. (2004) provide a strong support for application of an allometric scaling exponent that corresponds to caloric demand.

Price et al. (2008) reanalysed the toxicity dataset of 64 anti-neoplastic drugs evaluated previously by Schneider et al. (2004) and obtained very similar results (see Table 4-4).

Bokkers and Slob (2007) compared the ratios of NOAELs and BMDs of effects in mice with those of the same effects in rats. They calculated the GM of this ratio distribution with several scaling exponents from 1 to 0.65. Comparisons based on doses per kg body weight and day (exponent 1) yielded a GM of 2.01 (GSD 3.44) for mouse to rat ratios based on NOAELs and a GM of 1.81 (GSD 2.32) for ratios based on benchmark doses. For the latter the GM was closest to 1 with an exponent of 0.7. The slightly different ratios from NOAELs and BMDs are possibly the result of a different database (e.g. a NOAEL could not be determined for an effect, but calculation of a BMD is possible) and not of the differences of NOAEL and BMD-derived values per se, as the differences vanished if the authors limited the distributions only to effects which had a NOAEL and a BMD. The GSD of the distributions was 3.4 in the case of NOAELs and 2.3 in the case of BMDs, which is interpreted by the authors as a higher variability of ratios calculated from NOAELs due to the higher uncertainty of NOAEL values. Based on this observation they actually recommend to use higher assessment factors when a NOAEL is used as point of departure (see discussion below). The GSD of our analysis in this report with ratios from NOAELs was quite comparable (3.59 for oral exposures).

Escher et al. (2013) derived ratios from studies in the RepDose database where a comparable study result was available for the two species rat and mouse. The data was analysed separately for the routes inhalation, oral via gavage and oral via feed/water. Studies with inhalation exposure were further split into local and systemic effects. Because data-rich substances allow for more than one comparison, two datasets were generated. One dataset contained all possible comparisons and another contained only the comparison of medians of NOAELs from several studies per substance and species, i.e. only one value per substance. The GMs of these two datasets are quite comparable. The dataset with all comparisons is only about 50% bigger. In addition, the dataset based on medians is comparable to our analysis, therefore in the following the discussion is focused on this reduced dataset. In agreement with the concept of caloric demand scaling, all exposure routes which do not call for allometric scaling (inhalation exposure, food and drinking water studies with comparison based on concentrations in media) have a GM at 1.00. No difference between ratios for systemic and local effects after inhalation was observed. Studies performed with gavage were evaluated with (i.e. division of the GM by a factor of 1.78, according to caloric demand scaling) and without allometric correction. Without correction, the GM was 1.97, whereas the corrected GM was 1.13. These study results

are also in agreement with caloric demand scaling. GSD between 1.57 and 3.53 were obtained for the various datasets.
Kratchman et al. (2018) derived benchmark doses for a set of 41 NTP study reports (chronic and sub-chronic studies) per sex and species. They found that in these studies rats were significantly more sensitive than mice (comparison based on BMDL values calculated, in mg/kg bw/day). However, the authors did not calculate ratios from the BMDL values and, as only the lowest BMDL per study (either from rats or from mice) is reported in the publication, such a calculation cannot be performed from the data available. Therefore, the study is not included in the table below.

Table 4-4 Overview on empirical evaluations of species differences in toxicity over the last two decades

Reference	**Substances / dataset**	**Study characteristics**	**GM / medians**	**GSD**	**Dose descriptor for ratios**
Schneider et al. (2004)	Anti-neoplastic agents; n=63	5-days, parenteral	Medians: m/h: 8.0 (exp.: 7.4)[1] hamster/h: 7.6 (exp.: 5.9)[1] r/h: 2.6 (exp.: 4.9)[1] monkey/h: 2.4 (exp.: 2.2)[1] dog/h: 1.2 (exp.: 1.7)[1]	3.23 (combined dataset)	MTD
Schneider et al. (2004)	Pesticides; n=216	Subchronic studies	Medians: m/r: 2.22 (exp.: 1.9)[1] r/dog: 1.7 (exp.: 2.5)[1] m/dog: 6.0 (exp.: 4.5)[1]	-	MTD

Reference	Substances / dataset	Study characteristics	GM / medians	GSD	Dose descriptor for ratios
Price et al. (2008)	Antineoplastic agents; n=64	5-days, parenteral	Medians: m/h: 7.7 r/h: 3.0 monkey/h: 2.5 dog/h: 1.0		MTD
Bokkers and Slob (2007)	NTP studies; n=58	13-w and 2-y NTP studies; 228 datasets	GM: m/r: 2.01 (exp.: 1.7)[3]	3.44	NOAEL
Bokkers and Slob (2007)	NTP studies; n=58	13-w and 2-y NTP studies; 368 datasets	GM: m/r: 1.81 (exp.: 1.7)[3]	2.32	BMD
Escher et al. (2013)	RepDose® database; n=50 resp. 58	Repeated dose oral toxicity studies	GM, all m/r: Food/dw: 1.0 Gavage: 1.97 (exp.: 1.7)[3]	1.99 1.57	NOAEL
Escher et al. (2013)	RepDose® database; n=40 resp. 58	Repeated dose inhalation toxicity studies	GM, all m/r: systemic: 1.0 local: 1.0	3.53 2.51	NOAEL
Our evaluation	NTP studies; n=927	14-d, 13-w and 2-y oral NTP studies	GM, m/r: [2] 2.5 (exp.: 1.7)[3]	3.78	NOAEL

Reference	Substances / dataset	Study characteristics	GM / medians	GSD	Dose descriptor for ratios
Our evaluation	NTP studies; n=333	14-d, 13-w and 2-y in-halation NTP studies	GM, m/r: [2] 1.04 (exp.: 1)	3.61	NOAEL
Our evaluation	REACH studies; n=135	oral repeated dose studies	GM, m/r: [2] 1.52 (exp.: 1.7)[3]	3.85	NOAEL
Our evaluation	REACH studies; n=105	inhalation repeated dose studies	GM, m/r: [2] 0.92 (exp.: 1)	2.98	NOAEL
Our evaluation	REACH studies; n=72	oral repeated dose studies	GM, r/dog: [2] 1.47 (exp.: 2.9)[3]	4.71	NOAEL

Abbreviations: r: rat, m: mouse, h: human, exp.: value expected according to caloric demand scaling; dw: drinking water
[1] Expected values as reported in source
[2] For allowing comparison with other studies, for GM the inverse ratios are reported here (mouse vs rat; rat vs dog) compared to Table 3-1, Table 3-8 and Table 3-9.
[3] Calculated with body weights as given in Table 4-3

Our own evaluation of NTP and REACH data are in line with the existing evaluations, although the majority of data points belong to rat/mouse comparisons, two species with a rather small difference in body size.

The comparison of rat versus dog studies from REACH data is biased to some extent, as typically 2 years studies for rats are compared to 1 – 2 years studies with dogs, although for the latter species this comprises only a smaller portion of their life span. Under comparable conditions the ratio is expected to be higher.

4.5.2.2 Conclusions regarding applicability of allometric scaling rules

Overall, despite slightly different methodological approaches, the existing empirical evaluations provide a uniform picture:

- Larger species appear more sensitive than smaller ones, if doses are calculated on a body weight basis (mg substance/kg bw/day).
- If species comparisons are made based on exposure media (air concentration in case of inhalation studies, concentration in feed or drinking water, (see Escher et al. 2013), then species show, on average, similar sensitivities, in agreement with caloric demand scaling.
- All individual evaluations have strengths and weaknesses:
    - Our new evaluation of NTP studies is the largest reported so far, with data from 241 substances and 1260 ratios, but is restricted to data from mice and rats; also, the REACH data provide information for larger species to a very limited extent only.
    - Schneider et al. (2004) and Price et al. (2008) evaluated data from anti-neoplastic agents, which included many different species including humans; however, the endpoints evaluated differ slightly per species and the data represent 5-day exposures only.
    - Bokkers and Slob (2007) evaluated NTP data by calculating both NOAEL and BMD ratios, but the number of studies is limited and only mouse and rat data are available for these studies.
- Although there is some variability in the data, all these evaluations are in agreement with caloric demand scaling (most exponents derived for oral data are in a range from 0.65 to 0.8) (expected exponents: body weight scaling: 1; caloric demand scaling 0.75; body surface scaling 0.67)

It is therefore concluded that the current practice of applying caloric demand scaling in various regulatory frameworks (e.g. REACH, the German system for deriving OEL values) is supported by the empirical data.

4.5.2.3 Conclusions regarding remaining interspecies variability

Applying appropriate scaling factors is expected to result in a distribution with a geometric mean of 1, i.e. after scaling on average humans and experimental animals are of the same susceptibility. Nevertheless, on a substance-by-substance basis humans may show higher or lower susceptibility.

This variability observed in the datasets evaluated comprises several elements

- The substance-to-substance variability in species differences
- The uncertainty associated with the values used for calculating the ratios (BMD or NOAEL values).

The variability in our evaluation and the published data presented here are expressed in the form of GSD values. For a detailed explanation on the use of GSD to describe the spread of a distribution the reader is referred to the report on “Intraspecies

extrapolation" in the same report series of this research project and to the WHO/IPCS report on uncertainties in hazard characterisation (WHO 2014)[19]. GSD values for the NTP data in our evaluation were in the range of 3.6 – 3.8. We obtained a slightly lower GSD in the previous evaluation of anti-neoplastic agents (GSD = 3.23), as reported by Schneider et al. (2004). Bokkers and Slob (2007) obtained a similar GSD (3.44) for their evaluation of a smaller set of NTP studies, when calculating ratios from NOAEL values, but yielded a lower GSD (2.32) for ratios based on BMD values. This indicates that the inherent uncertainties in NOAELs might add substantially to the variability in interspecies ratios. When the authors tried to further correct the GSD for uncertainties in the BMD values (by assuming constant estimation errors), GSD was further reduced to approx. 2 (GSD = 1.98).

Based on their observations the authors proposed to apply different interspecies assessment factors if the point of departure is a NOAEL or a BMDL (Bokkers and Slob 2007). However, the uncertainty in interspecies ratios resulting from the uncertainty in the underlying database used for their calculation is different from the uncertainty of a certain POD (being either a NOAEL or a BMD) in a substance-specific context. We recommend addressing the uncertainty of the POD as a separate, mandatory step when deriving health-based guidance values. This can be done quite easily with benchmark doses by using the BMDL instead of the BMD but requires further assumptions in the case of a NOAEL. This will be addressed in a further report ("Synthesis report: Modelling of distributions of assessment factors, comparison with current methods and discussion of protection goals") in this project.

For characterising the substance-to-substance variability in interspecies extrapolation it would be preferable to eliminate the uncertainty of the underlying data. The evaluation by Bokkers and Slob (2007) is the only one allowing to assess the uncertainty resulting from using NOAEL ratios (unfortunately, the work by Kratchman et al. (2018) does not allow to calculate ratios in order to check the results by Bokkers and Slob (2007)), but the conclusion that BMD values are associated with less uncertainty and, hence, the obtained distribution is less variable, is theoretically justifiable. The BMD ratios still might include slight uncertainties coming with the BMD determination. On the other hand, the dataset evaluated contains a limited number of NTP studies and might underestimate the variability in larger and more heterogeneous datasets. In conclusion, deriving a concept, which includes application of caloric demand scaling and a distribution (with GM=1) accounting for uncertainty due to substance-to-substance variation is in agreement with the available data. A distribution derived from BMD ratios is expected to contain less uncertainty than one derived from NOAEL ratios.

In the IPCS report on uncertainty in hazard assessment (WHO 2014) it was suggested to create an own distribution for describing the uncertainty in determining the adequate

[19] WHO (2014): "GSD is a measure for the spread of a distribution, which is preferred over the standard deviation in case of lognormal or other right-sided distributions". It can be used to compare the spread of distributions. Note that GSD has no statistical definition in a sense that it represents the standard deviation of lognormally distributed values.

allometric scaling exponent. In the report a distribution is proposed representing a range of exponents from 0.66 to 0.74, to be used in addition to the distribution representing the remaining substance-to-substance species variability due to toxicokinetic and –dynamic differences. However, limitations in determining the “true” exponent in empirical investigations is also caused by substance-to-substance variability and deriving two distributions from the same empirical evaluation would lead to considering the same source of uncertainty twice. In consequence, we propose that he allometric scaling factors are considered as a correction factor when doses are expressed as amount per kg bodyweight and the remaining differences in toxicokinetics and toxicodynamics should be accounted for by a separate distribution.

4.6 Interspecies extrapolation for local effects in the respiratory tract

4.6.1 Dosimetry modelling

In a separate report in this project (“Human Equivalent Concentration and kinetic modelling of aerosols in the respiratory tract”) we discussed methods for considering species differences in the dosimetry of particulate matter in the lower respiratory tract. The so-called HEC (Human Equivalent Concentration) procedure uses the MPPD deposition model and considers differences in deposition and clearance of aerosols between rodents and humans. The conclusion of this report is that several open questions remain in the application of the HEC procedure as well as relevant uncertainties. As the difference between the starting concentration C (determined in a rodent inhalation study) and HEC is typically well below a factor of 10 and the remaining uncertainties are higher in most cases, the uncertainty in this method needs to be addressed when used for deriving OELs. It must be noted that the HEC procedure accounts for differences between species for deposition and clearance only. Toxicodynamic differences are not accounted for. Together with the high inherent uncertainty of the HEC procedure (which cannot be quantified currently) we therefore propose to use the same distribution as above to account for uncertainties in interspecies extrapolation in case of particulates assessed by the HEC procedure.

The target site of gaseous substances is largely determined by their water solubility and reactivity. Based on these characteristics US EPA established three categories of substances: Category 1 comprises substances with high water solubility and/or high reactivity (e.g. formaldehyde). These are deposited mainly in the upper respiratory tract. Category 3 gases are insoluble and of low reactivity (e.g. trichloroethylene). They reach the pulmonary region and can be absorbed there into the circulation. Category 2 (e.g. ozone) are in between (Kuempel et al. 2015; US EPA 1994). Due to anatomical and physiological differences in the nose interspecies dosimetry of category 1 and 2 gases is especially demanding. So-called computational fluid dynamics models take anatomical and air flow specificities of rats and humans into account and aim at predicting the local doses in the various parts of the respiratory tract. However, such

elaborate models exist for few substances only (Kalberlah et al. 2002; Kuempel et al., 2015).

4.6.2 Empirical studies

Empirical evaluations on species differences for respiratory toxicants are scarce. In our evaluation of NTP studies we did not find significant differences in rat to mouse ratios between local and systemic effects. No differentiation was made between effects in the upper or lower respiratory tract. Escher et al. (2013) found a GM of 1 for differences between rats and mice in case of inhalation studies of substances with local effects in the respiratory tract (again, without differentiation between upper and lower respiratory tract). The same result was obtained for systemic effects in inhalation studies. The GSD observed was 2.5 – 2.7, similar to the variability observed for systemic effects.

In a research project for BAuA potential differences between respiratory effects in humans and experimental animals were analysed in detail (Kalberlah et al. 2002; Kalberlah et al. 1999). The authors compared the available data for humans and experimental animals for eight data-rich respiratory irritants (acrolein, ammonia, chlorine, formaldehyde, sulfur dioxide, epichlorohydrin, ozone, and nitrogen oxides). All of them but nitrogen oxides exert their effects partly or predominantly in the upper respiratory tract. The authors concluded that differences in exposure conditions and types of endpoints investigated limit the comparability of NOAECs and LOAECs between humans and experimental animals. According to their conclusions the small set of substances evaluated point to a “marginally higher sensitivity on the part of humans on average”.

Brüning et al. (2014) developed a regulatory framework for assessing local effects in the upper respiratory tract, including sensory irritation effects transmitted via neuronal sensors. They analysed and compared data for a set of substances causing irritating effects in the upper respiratory tract: ethyl acrylate, formaldehyde, methyl methacrylate, acetaldehyde, ammonia, n-butyl acetate, hydrogen sulfide, and 2-ethylhexanol and concluded that an interspecies factor of 3 is sufficiently conservative to extrapolate from chronic irritating effects in experimental animals to a human NOAEC for sensory irritation. For substances with effects predominantly on the olfactory epithelium “it should be considered to reduce the default factor to 2”, as the airflow along the rat olfactory epithelium is higher, resulting in an approximately twofold high tissue burden in rats compared to humans.

4.7 New approaches – genomics etc

Newly emerging techniques might allow for alternative or supplementing approaches for interspecies extrapolation. Using toxicogenomics profiling (Kienhuis et al. 2009) compared biochemical pathways in vitro in rat and human hepatocytes and in rats in vivo for acetaminophen-induced liver toxicity. In a parallelogram approach using these differential genomics profiles they identified six pathways relevant for all three systems

and concluded that these pathways are relevant mechanistic indicators for the human in vivo situation.

In a more generic way Burgess-Herbert and Euling (2013) review existing comparative genomics approaches for addressing interspecies extrapolation. They describe advantages and disadvantages of analysing species differences at the gene or protein level, by comparing toxicity pathways or by establishing complex network descriptions of the biological systems in rats and humans. Their current potential contribution to risk assessment can be seen in substance-specific or pathway-specific predictions of species differences, e.g. by comparing effects of anti-androgens between species and develop a prediction model for substances of the same mode of action. The authors see a major challenge in using these methods for producing generic approaches for predicting quantitative differences between species.

5 Conclusions

In conclusion we recommend to

- Apply the concept of caloric demand (or metabolic rate) scaling in all cases for which no substance-specific assessment of interspecies differences is possible.
- This implicates that allometric correction factors (see Table 4-3) are used when (oral) doses are expressed as amount per kg bodyweight.
- This applies typically to repeated-dose studies with oral administration used for deriving OELs.
- No allometric factor is required if the OEL derivation starts from an air concentration; no differences were observed in that regard between substances exerting systemic versus those exerting local effects.
- Although the conclusions by Rhomberg and Wolff (1998) are based on a data evaluation bias the situation is less clear for short-term effects, which might be governed by peak concentrations rather than integrated system exposure.
- The concept of caloric demand scaling applies to substances with the toxic moiety being the parent compound or a metabolite (it was hypothesized that spontaneously deactivated metabolites follow other rules; however, no empirical data exist to prove this and it can be assumed that in these cases the delivery of the metabolite to the target organ is governed by the distribution of the parent compound, which is following allometric scaling rules).
- For considering the remaining uncertainty (after applying allometric scaling factors) due to substance-to-substance variability in toxicokinetics and toxicodynamics we propose to use a distribution centred around 1 (GM=1).
- Due to its higher quality, the empirical dataset derived from NTP data is preferable over the REACH dataset. However, both datasets include additional uncertainty from using NOAEL ratios. The empirical data obtained from the NTP data will be used to derive a distribution, which will be compared to the results of Bokkers and Slob (2007) to conclude on the best way to describe the remaining uncertainty after scaling. If possible, the uncertainty introduced by errors in using NOAELs for calculating ratios should be corrected, e.g. by adjusting the distribution according to this additional uncertainty or by directly using a distribution derived from BMDs.
- Due to the high inherent uncertainty of the HEC procedure (see separate report “Human Equivalent Concentration and kinetic modelling of aerosols in the lower respiratory tract”) it is proposed to use the same distribution as for systemic effects to account for uncertainties in interspecies extrapolation in case of particulates assessed by the HEC procedure.

References

Bokkers, B.G.H. (2009). Extrapolation in human health hazard characterization: a probabilistic approach, IRAS, Utrecht University, 2009.

Bokkers, B.G.H., Slob, W. (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. Critical Reviews in Toxicology 37, 355-373.

Boxenbaum, H. (1982). Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics. Journal of Pharmacokinetics and Biopharmaceutics 10, 201-227.

Boxenbaum, H. (1984). Interspecies pharmacokinetic scaling and the evolutionary-comparative paradigm. Drug Metabolism Reviews 15, 1071-1121.

Brüning, T., Bartsch, R., Bolt, H.M., Desel, H., Drexler, H., Gundert-Remy, U., Hartwig, A., Jäckh, R., Leibold, E., Pallapies, D., Rettenmeier, A.W., Schlüter, G., Stropp, G., Sucker, K., Triebig, G., Westphal, G., van Thriel, C. (2014). Sensory irritation as a basis for setting occupational exposure limits. Archives of Toxicology 88, 1855-1879.

Burgess-Herbert, S.L., Euling, S.Y. (2013). Use of comparative genomics approaches to characterize interspecies differences in response to environmental chemicals: challenges, opportunities, and research needs. Toxicology and Applied Pharmacology 271, 372-385.

Burzala-Kowalczyk, L., Jongbloed, G. (2011). Allometric scaling: analysis of LD_{50} data. Risk Analysis 31, 523-532.

Cao, Q., Yu, J., Connell, D. (2014). New allometric scaling relationships and applications for dose and toxicity extrapolation. International Journal of Toxicology 33, 482-489.

ECHA, European Chemicals Agency. (2012). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1, November 2012, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

Escher, S.E., Batke, M., Hoffmann-Doerr, S., Messinger, H., Mangelsdorf, I. (2013). Interspecies extrapolation based on the RepDose database - A probabilistic approach. Toxicology Letters 218, 159-165.

FDA, Food and Drug Administration. (2005). Guidance for Industry. Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers U.S. Department of Health and Human Services. Food and Drug

Administration. Center for Drug Evaluation and Research (CDER). https://www.fda.gov/media/72309/download.

Griem, P., Hassauer, M., Kalberlah, F., Oltmanns, J., Scheibner, J., Schneider, K., Schuhmacher-Wolz, U. (2002). Quantitative Differences in Xenobiotic Metabolism Between Experimental Animals and Humans, Federal Institute for Occupational Safety and Health. Dortmund/Berlin/Dresden.

Hu, T.M., Hayton, W.L. (2001). Allometric scaling of xenobiotic clearance: uncertainty versus universality. AAPS PharmSciTech 3, E29ff.

Huang, Q., Gehring, R., Tell, L.A., Li, M., Riviere, J.E. (2015). Interspecies allometric meta-analysis of the comparative pharmacokinetics of 85 drugs across veterinary and laboratory animal species. Journal of Veterinary Pharmacology and Therapeutics 38, 214-226.

Kalberlah, F., Föst, U., Schneider, K. (2002). Time extrapolation and interspecies extrapolation for locally acting substances in case of limited toxicological data. Annals of Occupational Hygiene 46, 175-185.

Kalberlah, F., Schneider, K., Schuhmacher, U.S., Voss, J.-U. (1999). Zeitextrapolation und Interspeziesextrapolation bei lokal wirksamen Stoffen mit begrenzter Datenlage, Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Fb 862 Wirtschaftsverlag NW Bremerhaven.

Kenyon, E.M. (2012). Interspecies extrapolation. In: B. Reisfeld, A. Mayeno (Eds), Computational Toxicology. Methods in Molecular Biology (Methods and Protocols), Vol. 929, , pp. 501-520.

Kienhuis, A.S., van de Poll, M.C., Wortelboer, H., van Herwijnen, M., Gottschalk, R., Dejong, C.H., Boorsma, A., Paules, R.S., Kleinjans, J.C., Stierum, R.H., van Delft, J.H. (2009). Parallelogram approach using rat-human in vitro and rat in vivo toxicogenomics predicts acetaminophen-induced hepatotoxicity in humans. Toxicological Sciences 107, 544-552.

Kratchman, J., Wang, B., Gray, G. (2018). Which is most sensitive? Assessing responses of mice and rats in toxicity bioassays. Journal of Toxicology and Environmental Health. Part A 81, 173-183.

Kuempel, E.D., Sweeney, L.M., Morris, J.B., Jarabek, A.M. (2015). Advances in inhalation dosimetry models and methods for occupational risk assessment and exposure limit derivation. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S18-40.

Mahmood, I. (2012). Interspecies scaling of biliary excreted drugs: prediction of human clearance and volume of distribution. Drug Metabolism and Drug Interactions 27, 157-164.

Mumtaz, M.M., Pohl, H.R. (2012). Interspecies uncertainty in molecular responses and toxicity of mixtures. EXS 101, 361-379.

Poulin, P., Arnett, R. (2017). Integration of a plasma protein binding factor to the Chemical-Specific Adjustment Factor (CSAF) for facilitating the estimation of uncertainties in interspecies extrapolations when deriving health-based exposure limits for active pharmaceutical ingredients: Investigation of recent drug datasets. Regulatory Toxicology and Pharmacology 91, 142-150.

Price, P.S., Keenan, R.E., Swartout, J.C. (2008). Characterizing interspecies uncertainty using data from studies of anti-neoplastic agents in animals and humans. Toxicology and Applied Pharmacology 233, 64-70.

Rhomberg, L.R., Wolff, S.K. (1998). Empirical scaling of single oral lethal doses across mammalian species based on a large database. Risk Analysis 18, 741-753.

Schneider, K., Oltmanns, J., Hassauer, M. (2004). Allometric principles for interspecies extrapolation in toxicological risk assessment - empirical investigations. Regulatory Toxicology and Pharmacology 39, 334-347.

Sharma, V., McNeill, J.H. (2009). To scale or not to scale: the principles of dose extrapolation. British Journal of Pharmacology 157, 907-921.

Travis, C.C. (1990). Tissue dosimetry for reactive metabolites. Risk Analysis 10, 317-321.

Travis, C.C., White, R.K. (1988). Interspecific scaling of toxicity data. Risk Analysis 8, 119-125.

Travis, C.C., White, R.K., Ward, R.C. (1990). Interspecies extrapolation of pharmacokinetics. Journal of Theoretical Biology 142, 285-304.

US EPA, Environmental Protection Agency. (1994). Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry, U.S. Environmental Protection Agency, Research Triangle Park NC. Office of Health and Environmental Assessment.

US EPA, Environmental Protection Agency. (2011). Recommended Use of Body Weight$^{3/4}$ as the Default Method in Derivation of the Oral Reference Dose. EPA/100/R11/0001. Final, Office of the Science Advisor. Risk Assessment Forum. Washington, DC

WHO, World Health Organization. (2009). Environmental Health Criteria 240, Principles and Methods for the Risk Assessment of Chemicals in Food, IPCS International Programme on Chemical Safety; World Health Organization, Geneva, Switzerland.

WHO, World Health Organization. (2014). Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. IPCS Harmonization Project Document 11, WHO, Geneva, Switzerland. http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf.

List of Tables

List of Figures

REPORT 8:
Intraspecies extrapolation

Research Project F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Summary

Consideration of inter-individual differences in susceptibility to chemical substances is a key aspect when deriving health-based guidance values. Such variability may have its origin in differences in toxicodynamics (i.e., inter-individual variation in responses of the target tissue to the same internal exposure) or differences in toxicokinetics (i.e., inter-individual variation in internal dose at the same external exposure). Various conditions are known to influence susceptibility, among them age, sex, genetics (e.g., polymorphisms of xenobiotic metabolising enzymes), epigenetic differences, and impaired health. Accordingly, quantification of the inter-individual ("intraspecies") variability for risk assessment purposes remains a challenge. Currently, methodologies for deriving OEL values use poorly justified default values.

Here we report results from the compilation and evaluation of a newly established database of human studies with

- 78 datasets (68 of which could be evaluated quantitatively) on differences in toxicokinetics
- 25 datasets on differences in toxicodynamics.

Variability in toxicokinetic data were characterised by log GSD values (the standard deviation of the logarithmised data). The median of log GSD values of the whole dataset was 0.146, equivalent to a factor of approx. 1.7 between the median and the 95th percentile of the population. The 95th percentile of log GSD of 0.355 corresponds to a factor of 3.8 to cover 95% of the population (the concept of log GSD for describing variability is further explained in Annex 2 of this report). A significant difference between data from oral and inhalation exposure was observed, with lower variability for inhalation data.

The data on toxicodynamics are associated with large uncertainties. For the difference between the lowest dose or concentration showing effects in some individuals and the highest dose or concentration showing no effects in others, a range from 3 to 201 was observed.

These results were compared and evaluated with existing evaluations in the literature. Substance-specific data on toxicokinetic differences, as well as case studies using PBPK modelling, result in toxicokinetic extrapolation factors in the range of 1.5 to 6, but higher factors are required for substances metabolised via polymorphic enzymes such as CYP2C9. A high agreement was seen between the Hattis database on toxicokinetic differences and our data on oral exposures.

Recently, a database on toxicodynamic variability was published, using high-throughput screening data of immortalised lymphoblastoid cells from over 1000 individuals representing different populations from five different continents. The variability in the toxic responses observed in vitro in these cell lines to 179 chemicals can be used to derive a distribution for toxicodynamics, which is largely in agreement with human in vivo data from the Hattis database.

In conclusion, a new database was compiled, which can be used in combination with published data to establish data-based distributions for toxicokinetic and toxicodynamic differences in susceptibility to chemical substances in the human adult population. We propose to use our new database for toxicokinetic differences and the in vitro dataset of Abdo et al. (2015) on toxicodynamic differences for OEL derivation.

Abbreviations

AUC	Area under the curve
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
CI	Confidence Interval
CV	Coefficient of variation
C_{max}	Peak concentration
COPD	Chronic obstructive pulmonary disease
CSAF	Substance-specific adjustment factors
CYP	Cytochromes P450 superfamily of enzymes
ECHA	European Chemicals Agency
EF/AF	Extrapolation factor /Assessment factor
EPA	Environmental Protection Agency (in the US)
FEV_1	Forced expiratory volume in one second
GM	Geometric mean
GSD	Geometric standard deviation
ICRP	International Commission on Radiological Protection
IPCS	International Programme on Chemical Safety
IVIVE	In vitro – in vivo extrapolation
NOAEC	No observed adverse effect concentration

NTP	National Toxicology Program
OEL	Occupational exposure limits
PBPK	Physiologically based pharmacokinetic modelling
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals, Regulation (EC) No 1907/2006
SD	Standard deviation
TD	toxicodynamic
TK	toxicokinetic
WHO	World Health Organisation

1 Introduction and regulatory background

Differences in the susceptibility to toxicants between individuals can have multiple reasons (Falk-Filipsson et al. 2007; Nebert 2005). They can be caused by differences in toxicodynamics, i.e., variation in responses of the target tissue to the same internal exposure or by differences in toxicokinetics. The latter can occur at various levels: absorption, distribution, metabolism or excretion of the substance. Variations in the type and quantity of metabolites can be the result of genetic polymorphisms of xenobiotic-metabolising enzymes, such as CYP2C9 or CYP2D6 (Dorne 2007; Gentry et al. 2002). However, differences at the level of enzyme activity are not necessarily resulting in similar differences in internal doses. Only if key steps controlling the internal dose of the critical agent are impacted, a large quantitative effect can be expected (Bois et al. 2010).

Susceptibility can be influenced by age, sex, genetics (e.g., polymorphisms), epigenetic differences, impaired health and other factors (chapter 3.1). Due to these many reasons, quantification of the inter-individual (“intraspecies”) variability for risk assessment purposes remains a challenge. For the purpose of deriving health-based guidance values for the general population, WHO proposed to split the intraspecies extrapolation factor (usually of the magnitude 10) in two subfactors, for differences in toxicokinetics (factor 3.16) and toxicodynamics (factor 3.16) (WHO 1994), thus allowing to replace the subfactors by substance-specific adjustment factors (CSAF) in cases where substance-specific information is available and sufficient for quantifying the variability (Bhat et al. 2017; WHO 2005). Use of substance-specific data on inter-individual variability is also encouraged by the US Environmental Protection Agency (US EPA 2014) and in the respective ECHA guidance document for REACH (ECHA 2012). Default values of 5 for workers and 10 for the general population are proposed in the latter document, under the assumption that the variability in the general population, which includes children, elderly and diseased people, is higher than within workers. Approaches for considering inter-individual variability for deriving OELs vary considerably, as described in the separate project report “Comparison of methods for deriving OELs”.

In chapter 2 we present the methodological approach and the results of an evaluation of data from literature performed in this project. The evaluation comprises toxicokinetic and toxicodynamic data and tries to include industrial chemicals and inhalation exposures to the extent possible.

Existing quantitative evaluations of existing data are presented in chapter 3.2. Several authors used existing data on chemical substances (mostly pharmaceuticals) to derive CSAF for the toxicokinetic subfactor. Among them are the efforts by Dorne and colleagues to derive metabolism-pathway specific toxicokinetic assessment factors (Dorne 2007; Dorne et al. 2005) and the extensive database compiled by Dale Hattis and colleagues (Hattis et al. 2002; Hattis and Lynch 2007).

Zeise and colleagues summarise currently used approaches for individual, data-rich chemicals using physiology-based pharmacokinetic models (PBPK, in the following used in the same sense also for non-pharmaceuticals) (Zeise et al. 2013). Examples

are discussed in chapter 3.3. But these and other authors also outline how new techniques such as high-throughput screening with in vitro tests can be used to obtain information on inter-individual variability (Axelrad et al. 2019; Dornbos and LaPres 2018; Zeise et al. 2013). This is further explored in chapter 3.4.

In chapter 4 the results from the new data compilation and the available empirical data are discussed and proposals for distributions to be used for deriving OELs are developed.

2 Evaluation of literature data on inter-individual variability

2.1 Methodological approach

2.1.1 Literature search

Literature searches to identify publications with quantitative data on inter-individual differences were performed in summer 2018 in the PubMed database and were continued until October 2019. Only studies with humans were searched.

Details of the search strategies are documented in Annex 1.
Initially, searches were restricted to publications from the last 10 years. Screening of the search results revealed that data on toxicokinetics of pharmaceuticals after oral exposure were overrepresented. Therefore, additional searches without time restriction were performed in Pubmed with the objective to find more studies using industrial chemicals or inhalation exposure and to detect more data on differences in toxicodynamics.

2.1.2 Evaluation strategy

Abstracts of hits obtained from the data searches were screened for a high probability to find quantitative and relevant data in the publication. These publications were retrieved in original and evaluated for the two relevant endpoints “kinetic-” and “dynamic” effects in humans. Results were documented in a Microsoft Excel® table.

2.1.3 Pharmaco-/toxicokinetic effects

Studies were selected for evaluation if:

- route of substance application was oral or inhalation,
- relevant kinetic parameters like AUC or C_{max} were reported,
- the population studied did not mainly consist of children,
- statistical data coming from at least 4 individuals were reported,
- results were not only reported for a highly selective subgroup of individuals (e.g. population with a selected CYP polymorphism),
- data were available that allowed to characterise variability in the study group.

Studies relevant for evaluation were documented with author, year of publication, full citation and study characteristics (e.g. study in volunteers, placebo-controlled etc.). In addition, the name of the applied substance, the function (if known) and the substance class (pharmaceutical or industrial chemical) were reported. The group characteristic (with group size, further details on the group like sex or age, reported influence factors or state of health) were mentioned as well as application characteristics like route of application, frequency and dose.

Toxicokinetic parameters used were Area under the curve (AUC) of the plasma concentration-time curve and the maximum plasma concentration (C_{max}). In rare cases also excretion in urine was used. In order to be able to characterise variability either individual data or statistic values (mean, SD, GM, CV etc.) were retrieved as given in the publication and documented in Microsoft Excel®.

It was assumed that the variability in internal dose measures (e.g. AUC, C_{max}) reflects the variability in the external dose needed to achieve a definite internal dose. The underlying (simplistic) assumption is that for all individuals in the study group the relationship between internal and external exposure (in a certain, limited dose range) is linear, although not necessarily with the same slope for each individual.

The following example is intended to demonstrate how the studies were evaluated in the context of the current analysis:

Wenker et al. (2001) performed a study with 20 male volunteers between the age of 18 and 37. The subjects were exposed on separate occasions to 104.4 ± 3 or 360 ± 20 mg styrene/m³ for 1 h while performing physical exercise on a bicycle ergometer. Blood samples were taken up to 180 min after the start of exposure and urine was sampled up to 24 h after the end of exposure. The publication reports mean and SD, CV and the range of values for C_{max}, AUC, $t_{1/2}$ and the clearance for styrene and the metabolites mandelic acid and phenylglyoxylic acid. In addition, the publication aimed at identifying the influence of genetic polymorphisms.
For the evaluation, only one parameter for the main substance (styrene) was selected for documentation in the Microsoft Excel® table. The AUC for styrene obtained after exposure to the lower concentration (104 mg/m³) was selected as the relevant parameter to avoid any high dose phenomena under these short-term exposure conditions. For this parameter the mean ± SD, GM and min/max values are reported in the study.

Generally,
- AUC was given preference over C_{max} or other parameters
- If several concentrations were tested, the values reported for the lowest concentration were selected
- If single and repeated experiments were performed, results from repeated exposure were preferred, since they better represent exposure at workplaces.

All in all 74 entries for toxicokinetics were generated in the Microsoft Excel® Table.

For characterising the inter-individual variability within one dataset, the standard deviation of the logarithmised data was used. In analogy to previous evaluations (WHO 2014) this value is called log GSD (logarithm to base 10). The concept of log GSD is introduced and explained in detail in Annex 2.

For each evaluated study, a log GSD value was derived under the assumption that the reported distribution parameters were derived from lognormally distributed data. The calculation of the log GSD was performed with the formulae given in Table 2-1, which is ranked from top to bottom according to prioritisation of calculation, e.g. if a mean,

SD and CV is available, the calculation was performed according to the topmost formula, using the mean and SD. If only minimum and maximum values were given, no log GSD could be calculated, and the studies were not used. Table 2-1 shows how log GSD was calculated from various parameters given in the studies.

Table 2-1 Calculation of log GSD based on different statistical parameters given in the publications

Statistical parameters available	**Calculation of log GSD**
mean ± SD ($\mu_{lin} \pm \sigma_{lin}$ on the linear scale)	log GSD = $\log_{10}\left(e^{\sqrt{\ln\left(1+\frac{\sigma_{lin}^2}{\mu_{lin}^2}\right)}}\right)$
CV (CV [in %] on the linear scale)	log GSD = $\log_{10}\left(e^{\sqrt{\ln(1+CV^2)}}\right)$
GM[§], confidence interval, n ($GM, CI_{upper,\alpha,}, CI_{lower,\alpha}$ on the linear scale)	log GSD = $\log_{10}\left(e^{\left(\ln GM-\frac{\ln CI_{upper,\propto}}{qt^{\#}\left(1-\frac{\alpha}{2},\, n-1\right)}\right)\cdot\sqrt{n}}\right)$
Symmetric percentiles (e.g. 25th and 75th percentile, in which case α = 0.25) ($Quantile_{\alpha,}, Quantile_{1-\alpha}$ on the linear scale)	log GSD = $\log_{10}\left(e^{\frac{\ln Quantile_{1-\alpha}-\ln Quantile_{\alpha}}{2\,*\,qnorm(1-\alpha)^{\$}}}\right)$

[§] The GM is not necessary under assumption of lognormality. Yet, the GM was given in all cases where log GSD had to be derived from a CI, therefore the log GSD was derived using the distance from the expected value and one of the boundaries of the CI. The same calculation using the alternative boundary of the CI was used to scrutinize the assumption of lognormality

[#] qt(p, df) is the quantile of the t-distribution, where p is the probability (e.g. p = 0.975 for a 95% CI, in which case α = 0.05) and df are the degrees of freedom

[$] qnorm(p) is the quantile of the standard normal distribution with probability p

2.1.4 Pharmaco-/toxicodynamic effects

For evaluating differences in toxicodynamics, a linear relationship between external dose/concentration and effect measure cannot be assumed. Therefore, in order to obtain information on differences in doses/concentrations leading to a similar effect level, studies using a range of doses or concentrations were sought.

Studies were selected for evaluation if:

- the population studied consisted (predominantly or completely) of adults,
- individual effect data for at least two different doses/concentrations spread wide enough to observe the range of different susceptibilities were reported.

Studies with oral or inhalation exposure (“external exposure”) of the substance were considered to provide indications on inter-individual differences due to both toxicokinetic and toxicodynamic reasons. Human studies fulfilling the above criteria and using parenteral (non-inhalative) forms of applications (intravenous,

subcutaneous) were also included. In an approximation it was assumed that these datasets show predominantly differences in susceptibility due to toxicodynamic reasons (as variability due to differences in absorption in the gastrointestinal or respiratory tract is without impact in these cases).

As for pharmaco-/toxicokinetic effects, studies relevant for evaluation were documented reporting author, year of publication, full citation and study characteristics. In addition, the name of the applied substance, the function (if known) and the substance class (pharmaceutical, industrial chemical) were reported. The group characteristic (with group size, further details on the group like sex or age, reported influence factors or state of health) were mentioned as well as application characteristics like route of application, frequency and dose. Finally, the effect type as measured and documented in a study and dose ratios for the pharmaco/-toxicodynamic endpoint relevant in the individual studies were listed. For details on the evaluation see chapter 2.2. As an example, one study is described and the procedure how toxico-/pharmacodynamic information was evaluated is shown:

In a study by Chriguer et al. (2005) the sensitivity of individuals to glucocorticoids was evaluated in 40 healthy males and females (21 females, 19 males, 22 - 42 years). Plasma cortisol levels were measured after oral administration of 0.25, 0.5 or 1 mg dexamethasone. All volunteers randomly received all doses, at an interval of at least one week. For the evaluation, it was checked on an individual level if the same effect was observed to the same extent at different doses. In this case, plasma cortisol levels of two individuals in the highest exposure group showed comparable values to individuals in the 0.25 mg group. This indicates that in the selected collective a variability to the effect of dexamethasone on cortisol levels of **at least** 4 (1.0 / 0.25) must be acknowledged. For the evaluation a factor of 5 (next higher integer) was documented in the Microsoft Excel® table for this study.

For the quantitative evaluation ratios are calculated by dividing the highest dose or concentration without effects in some individuals by the lowest dose or concentration with effects, rounded to the next higher integer number. Using the next higher integer number should signal that the ranges of concentrations investigated were mostly not large enough to fully cover the differences.

2.2 Data evaluation

For both endpoints (pharmaco-/toxicokinetic and pharmaco-/toxicodynamic effects) the procedure of data evaluation is described in the following.

2.2.1 Pharmaco-/toxicokinetic data

2.2.1.1 Distribution of results from all evaluated studies

Seventy-four datasets were evaluated in detail and documented in the Excel file “Database inter-individual variability”. Sixty-eight datasets provided useful information. Six studies were dismissed, mainly because only ranges (maximum, minimum) were

reported. In Annex 3 a tabular summary of the individual datasets documented in the Excel® file is given.

The plots used for visualization and the statistical procedure to evaluate differences between distributions are described in more detail in the report on “Exposure duration extrapolation”. Briefly, each distribution is visualized by the combination of a density, a dot plot and a box plot and distributions were compared based on the 95% confidence intervals after bootstrapping.

The distribution of all available log GSD values is shown in **Figure 2-1**. The GM of the log GSD values is 0.14[1] and the 75% percentile is 0.22[2]. The most important parameters describing the resulting distribution are given in **Table 2-2**. The dataset is comprised of studies which differ in several factors in their study design. Some of these experimental factors are regarded as having a potential influence on the distributions. In the following, the dataset is split according to these factors and the resulting distributions are compared.

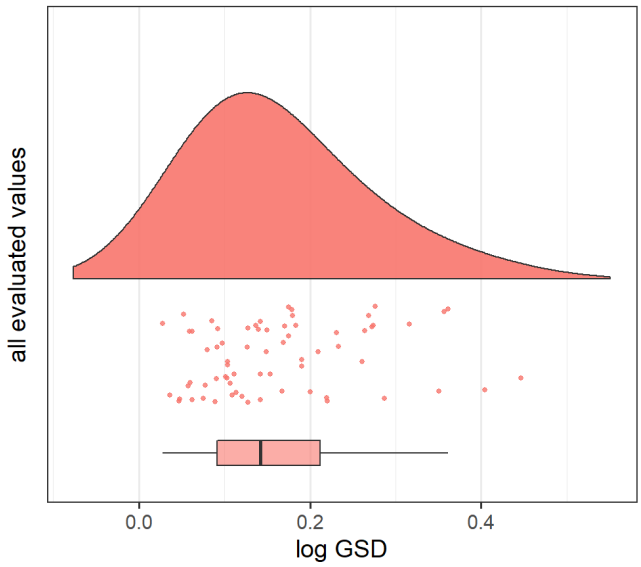


Figure 2-1 Distribution of the log GSD values from all evaluated studies for toxicokinetics intraspecies extrapolation.

[1] 95% CI of the GM: 0.12 – 0.16
[2] 95% CI of the 75th percentile: 0.18 – 0.27

Table 2-2 Parameters of the distribution of the log GSD values from all evaluated studies for toxicokinetics intraspecies extrapolation

Mean	SD	GM	GSD	5%	Median	75%	95%	n
0.166	0.094	0.141	1.830	0.049	0.146	0.220	0.355	68

2.2.1.2 Stratification by exposure route

The dataset was split into values from studies with oral and inhalation exposure. Four studies had an exposure route which does not fit into these categories (primarily applications with uptake via mucosae) and were not included in **Figure 2-2**. The exposure route had an influence on differences in toxicokinetics, as the GM for oral data was significantly higher compared to inhalation data (a difference was considered statistically significant when the confidence intervals didn't overlap at the 95th percent level)[3]. Yet when the distributions are compared on basis of their 75% percentile, the differences are not significant[4]. A summary of the distribution parameters is given in **Table 2-3**, which includes the distribution of values from studies not categorized as having oral or inhalation exposure.

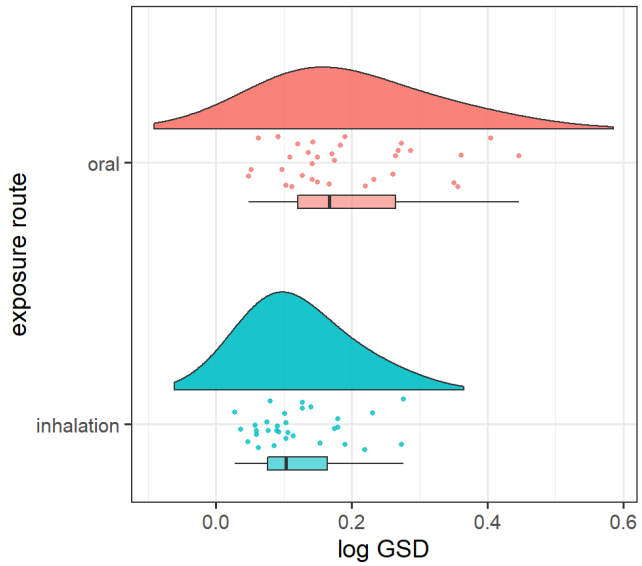


Figure 2-2 Distribution of the log GSD values from the studies for toxicokinetics intraspecies extrapolation, separated by exposure route.

[3] 95% CI of the GM, oral: 0.14 – 0.20; inhalation: 0.09 - 0.14
[4] 95% CI of the 75-percentile, oral: 0.18 – 0.35; inhalation: 0.13 - 0.22

Table 2-3 Parameters of the distribution of the log GSD values from the evaluated studies for toxicokinetics intraspecies extrapolation, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	0.194	0.104	0.168	1.764	0.058	0.167	0.264	0.379	33
inhalation	0.130	0.070	0.111	1.804	0.042	0.106	0.179	0.252	31
other	0.224	0.064	0.218	1.305	0.174	0.205	0.236	0.300	4

2.2.1.3 Stratification by substance class

Categorising studies by the nature of the investigated substance (industrial chemical or pharmaceutical compound) revealed no significant differences of the GM[5] and 75% percentile[6] between the compared groups. Still, it appears that industrial chemicals tend to have a smaller intraspecies variability in toxicokinetics according to our analysis (see **Figure 2-3**). A summary of the most important parameters of the two distributions given in **Table 2-4**.

The application route is closely linked to the type of substance: exposure to industrial chemicals was primarily (28 out of 31 datasets) via inhalation in the evaluated studies, while pharmaceuticals were primarily applied orally. Therefore, the differences by the substance class could well be just the mere correlation with the exposure path (or vice versa).

[5] 95% CI of the GM, industrial chemicals: 0.09 – 0.14, pharmaceuticals: 0.14 – 0.19
[6] 95% CI of the 75th percentile, industrial chemicals: 0.11 – 0.27, pharmaceuticals: 0.18 – 0.29

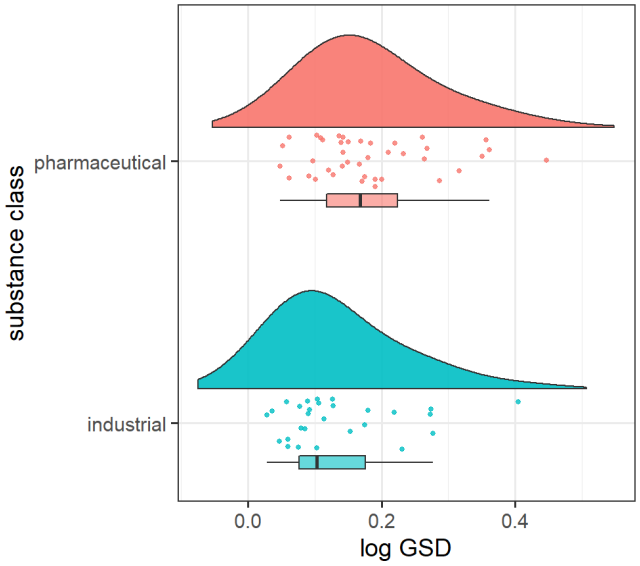


Figure 2-3 Distribution of the log GSD values from the evaluated studies for toxicokinetics intraspecies extrapolation, separated by substance class (industrial or pharmaceutical substance).

Table 2-4 Parameters of the distribution of the log GSD values from the evaluated studies for toxicokinetics intraspecies extrapolation, separated by substance class

Substance class	Mean	SD	GM	GSD	5%	Median	75%	95%	n
pharmaceutical	0.186	0.091	0.165	1.665	0.062	0.170	0.224	0.357	40
industrial	0.138	0.092	0.113	1.942	0.040	0.105	0.189	0.276	28

2.2.1.4 Stratification by health status of volunteers

The analysis whether the health status of the volunteers has an impact on the distributions is hampered by the relatively low number of evaluated studies with volunteers with impaired health. Based on the available data, no significant difference can be observed for the GM[7] and 75% percentile[8] of the log GSD values.

[7] 95% CI of the GM, healthy volunteers: 0.12 – 0.16, not healthy volunteers: 0.11 – 0.23
[8] 95% CI of the 75th percentile, healthy volunteers: 0.17 – 0.27, not healthy volunteers: 0.19 – 0.27

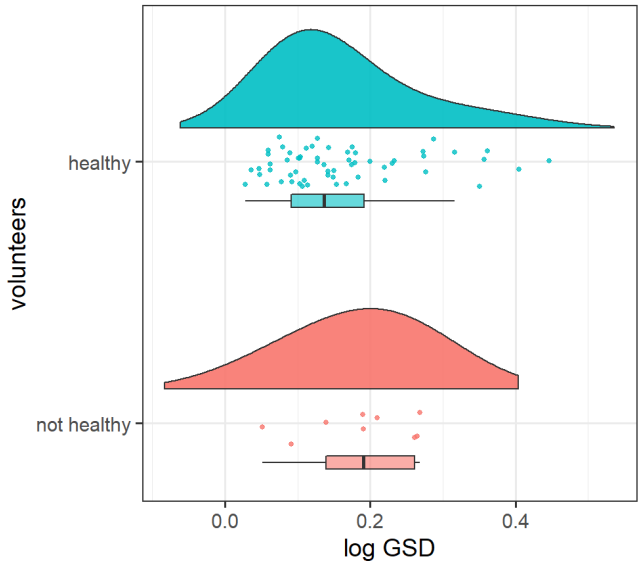


Figure 2-4 Distribution of the log GSD values from the evaluated studies for toxicokinetics intraspecies extrapolation separated by health status of the volunteers.

Table 2-5 Parameters of the distribution of the log GSD values from the evaluated studies for toxicokinetics intraspecies extrapolation, separated by health status of volunteers

Health status	**Mean**	**SD**	**GM**	**GSD**	**5%**	**Median**	**75%**	**95%**	**n**
not healthy	0.185	0.078	0.165	1.745	0.068	0.190	0.261	0.267	9
healthy	0.163	0.096	0.138	1.844	0.048	0.142	0.218	0.357	59

2.2.1.5 Discussion

Sixty-eight studies were evaluated, and log GSD values were derived. Splitting the data into values coming from oral or inhalation exposure showed that the exposure route had an influence on inter-individual variability of toxicokinetics; the GM obtained from oral data was significantly higher than the one from inhalation data. It was observed that the application route was closely linked to the substance class with pharmaceutical being normally applied orally and industrial chemical via inhalation. Therefore, the differences seen for the two exposure routes might be due to a correlation with the substance class (or vice versa).

2.2.2 Pharmaco-/toxicodynamic data

Twenty-five datasets were identified, 12 of them with inhalation exposure, 5 with oral and 8 with parenteral (6 x i.v., 2 x s.c.) administration. In general, quantification of the differences was difficult, as in several studies the dose or concentration range was not wide enough to be sure that the whole range of susceptibilities was included, which potentially might lead to an underestimation of differences. On the other hand, studies also included subjective symptom reporting, with the potential to overestimate differences in susceptibility. For example, in the study by Hine et al. (1960) volunteers were exposed to concentrations of triallylamine of 0.5 to 100 ppm and 4 individuals reported symptoms of eye and nose irritation at the lowest concentration, whereas 2 did not report symptoms even at 100 ppm, leading to a ratio of 201 in our evaluation. The obtained ratios range from 3 to 201.

Below we characterize the results by statistical parameters. Due to the high values in the graphical presentation $\log_{10}$ (dose ratios) were plotted. In the text and tabular summaries, the dose ratios are presented on the normal scale (i.e. without taking the $\log_{10}$). Again, in Annex 3 a tabular summary of the individual datasets documented in the Excel® file is given.

Note that the individual values (dose ratios) of each dataset represent the highest dose or concentration without effects in any individual divided by the lowest dose or concentration with effects, increased to the next higher integer. A factor derived from these data for covering the difference between average and high susceptibility would need to be lower (division by factor 2 in case of normally distributed data).

2.2.2.1 Distribution of results from all evaluated studies

The distribution of the derived dose ratios is centred around a GM of 7.37[9] and has a 75th percentile of 8.00[9]. Few extreme values are shifting arithmetic parameters strongly towards high values (**Figure 2-5**, **Table 2-6**). The ratios are derived from studies on pharmaceutical and industrial substances applied by different routes. The influence of these experimental factors is evaluated by stratifications in the next paragraph.

[9] 95% CI of the GM: 5.44 - 10.92 and of the 75th percentile: 6.25 – 16.00

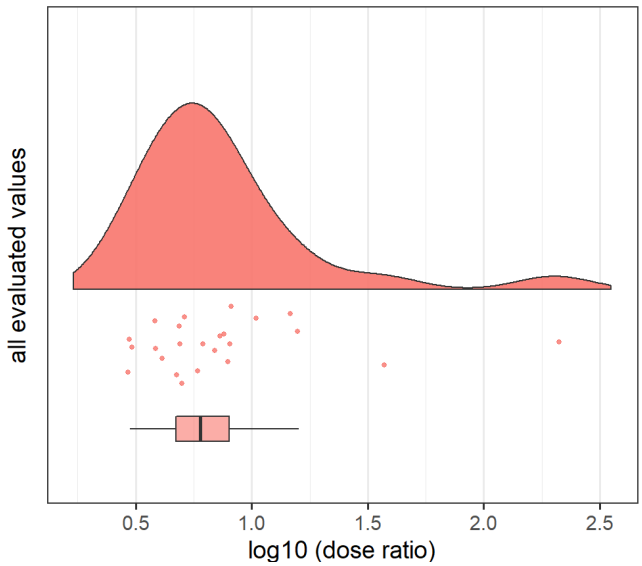


Figure 2-5 Distribution of the log (dose ratio) values from all evaluated studies for pharmaco-/toxicodynamic effects.

Table 2-6 Parameters of the distribution of the dose ratios from all evaluated studies for pharmaco-/toxicodynamic effects

Mean	SD	GM	GSD	5%	Median	75%	95%	n
15.92	40.01	7.39	2.48	3.00	6.00	8.00	33.00	24

2.2.2.2 Stratification by exposure route and substance class

The low number of ratios is making it difficult to evaluate the influence of the factors exposure route and substance class on variability of pharmaco-/toxicodynamic effects. For example, only 4 ratios are available for the oral path and only 5 ratios are available for parenteral (i.v. and subcutaneous administration). Geometric means indicate a lower variability for orally applied substances than for substances applied parenterally or via inhalation (**Table 2-7**), which is statistically significant even considering the low number of compared ratios[10]. In addition, the 75th percentile is significantly lower for orally applied substances than for the other paths[11]. However, it should be noted that these comparisons are severely impacted by the two higher values from the Hine et al. study with inhalation exposure.

[10] 95% CI of the GM for oral: 3.00 - 4.47, for inhalation: 5.30 – 17.69, for parenteral: 4.44 – 11.37
[11] 95% CI of the 75th percentile for oral: 3.00 – 5.00, for inhalation: 6.00 – 77.25, for parenteral: 5.00 – 14.00

Whether the substance is a pharmaceutical or an industrial chemical had a lower influence on the distribution (**Table 2-8**). The pharmaceutical substances primarily consist of orally or parenterally applied substances, while the industrial chemicals, in essence, are the ones which are tested via inhalation. In consequence, the lower GM of toxicodynamic variability of orally applied substances correlates with a lower point estimate of the variability of toxicodynamic effects of pharmaceuticals (GM for pharmaceuticals: 6.09, for industrial chemicals: 8.76). However, in this case the difference is not statistically significant[12]. The third category for substance class, "other" comprises only a single substance (a food constituent) and does not allow a meaningful comparison.

For 11 of the 12 industrial chemicals tested via inhalation the endpoint investigated was related to irritating effects, either in the upper or lower respiratory tract. Therefore, the large variation observed for industrial chemicals is more exactly attributed to "industrial chemicals causing respiratory irritation after inhalation".

Table 2-7 Parameters of the distribution of the dose ratios from the evaluated studies for pharmaco-/toxicodynamic effects, separated by exposure route

Route	Mean	SD	GM	GSD	5%	Median	75%	95%	n
oral	3.75	0.96	3.67	1.28	3.00	3.50	4.25	4.85	4
parenteral	8.20	4.44	7.13	1.86	3.40	8.00	11.00	13.40	5
inhalation	24.50	56.27	8.76	3.14	4.00	6.00	7.25	110.25	12
other	10.67	4.62	10.08	1.49	8.00	8.00	12.00	15.20	3

[12] 95% CI of the GM for pharmaceuticals: 4.29 – 8.69, for industrial chemicals: 5.50 – 17.77

Table 2-8 Parameters of the distribution of the dose ratios from the evaluated studies for pharmaco-/toxicodynamic effects, separated by substance class

Substance class	Mean	SD	GM	GSD	5%	Median	75%	95%	n
pharmaceutical	7.27	4.61	6.09	1.86	3.00	5.00	9.50	15.00	11
industrial	24.50	56.27	8.76	3.14	4.00	6.00	7.25	110.25	12
other	8.00	NA	8.000	NA	8.00	8.00	8.00	8.00	1

2.2.2.3 Discussion

The dataset on toxicodynamic differences contains various uncertainties:

- Only a limited number of datasets could be identified (n=25), as the condition was to identify human studies with a broad range of different exposure concentrations
- Some of the studies had only very limited number of participants per dose group and/or exposure was repeated at different exposure levels with the same individuals
- In some cases, individual data were reported only graphically and had to be extracted from figures
- Quantification of the differences between susceptible and less susceptible individuals was hampered by the limited exposure range in most studies, a possible reason for underestimating the ratios
- The identified studies include older ones with limited documentation and/or subjective reporting of symptoms
- All quantitative comparisons are severely impacted by the study by Hine et al. (1960), which resulted in ratios for two substances of 36 and 201; without these two values the arithmetic mean for industrial chemicals would be reduced from 24.5 to 5.7, and the geometric mean would decrease from 8.76 to 5.56.

Despite these uncertainties, it can be concluded that relevant differences between exposure routes were observed: differences in susceptibility were lower after oral versus inhalation exposure (both oral and inhalation exposure are taken to be indicative for differences due to toxicokinetic **and** -dynamic reasons) or parenteral applications (taken to be indicative for predominantly toxicodynamic reasons). The GM for datasets with parenteral administration was not lower than after oral exposure (indeed, it was higher). This indicates a relevant contribution of toxicodynamic reasons.

3 Literature evaluation

3.1 Reasons for inter-individual differences in susceptibility

3.1.1 Genetic disposition

Genetic variation can influence toxicity of substances in various ways. For example, anatomical differences may influence the airflow of vapours in the upper respiratory tract (Garcia et al. 2009). Biomolecules such as cellular transporters, nuclear receptors and enzymes show genetic variability in the human population (Kozyra et al. 2017; Lee and Ho 2017). Prominent examples are xenobiotic metabolising enzymes, which are expressed polymorphically, with pronounced consequences for the metabolism of chemicals (Haber et al. 2002; Thummel and Lin 2014). However, as mentioned earlier, polymorphisms only cause differences in the metabolism of a substance, if the specific metabolic reaction catalysed by the enzyme is becoming a rate limiting step. Polymorphic enzymes considered to be responsible for large differences in the metabolism of substances are for example phase I enzymes CYP2D6 (Dorne et al. 2005; Naumann et al. 2004), CYP2C9 (Dorne 2007; Dorne et al. 2005; Gentry et al. 2002), CYP2D19 (Dorne 2007; Dorne et al. 2005) or CYP2E1 (Neafsey et al. 2009) and phase II enzymes like glutathione S-transferase T1 (El-Masri et al. 1999; Jonsson and Johanson 2001) or N-acetyl transferase 2 (Dorne et al. 2005).

Genetic differences influencing chemical metabolism may also become evident between ethnic groups, as shown by Ning et al. (2017) for the metabolism of estragole to the ultimate carcinogenic metabolite 1’-sulfooxyestragol, which was higher in Caucasian compared to Chinese individuals.

3.1.2 Epigenetics

The term “epigenetics” comprises a variety of processes related to control of gene expression, among them covalent modifications of DNA (e.g. (de-)methylation), posttranslational modification of histones or reorganisation/repackaging of nucleus molecules, and regulation of gene expression by noncoding RNAs. Epigenetic alterations of expression of drug-metabolising enzymes or transporter proteins on the individual level may lead to differences in responses to drugs and other substances (Ivanov et al. 2012). Non-coding regulatory microRNA has been found to be involved in regulation of CYP enzymes and drug metabolism (Gomez and Ingelman-Sundberg 2009; Tracy et al. 2016).

3.1.3 Age

Several processes and capacities might change with increasing age, which can influence the disposition and metabolism of substances, e.g. changes in physiological functions and constitution (e.g. body fat, hydration status) leading to changes in disposition or changed absorption due to changes in gastric pH and functional changes

of the gastrointestinal tract, or decreased clearance due to decreases in blood flow and glomerular filtration rates (Thummel and Lin 2014).

In an extensive literature review, Clewell et al. (2002) gathered information on inter-individual differences in pharmacokinetics due to age or gender. Examples of higher and lower absorption in the gastrointestinal tract in the elderly compared to younger adults were found. Dermal absorption showed a decreasing tendency in the elderly. No clear differences were documented regarding metabolism, whereas glomerular filtration and tubular secretion in the kidney are generally decreased for lipophilic and hydrophilic substances.

Streeter and Faria (2017) evaluated toxicokinetic studies (Cmax, AUC) for 206 pharmaceutical substances involving groups of young adults and elderly people. Within group variation (in both the young and elderly) was low in most cases and the distance between means and mean + 2 standard deviations (assuming normal distributions) could be covered by a factor of 2 in most cases. However, for some substances the authors observed large differences between the two groups and, when calculating the distance between the mean for young adults and mean of elderly plus 2 standard deviations, then factors up to 7 (in one case 15.7) were obtained. This indicates that for individual substances age might have a large impact on toxicokinetic parameters.

3.1.4 Sex

Physiological differences between sexes might result in differences in toxicokinetic properties of substances (Gochfeld 2007, 2016). For example, differences in the composition of bile might influence absorption from the gastrointestinal tract (Nicolas et al. 2009). However, only few studies exist investigating the quantitative consequences for internal exposure. Chen et al. (2000) compared AUC and Cmax values for men and women from 26 studies on bioequivalence of drugs. They found that differences were mostly below a factor of 2. Few differences in toxicokinetics between sexes were identified in the review by Clewell et al. (2002). Due to a different lean body mass of women and men, differences in distribution of substances in the body can occur: the volume of distribution in women compared to men is higher for lipophilic substances and lower for hydrophilic substances. Differences were also observed regarding the metabolism: some CYP450 enzymes (CYP1A2, 3A4, 2D6) are reported to have lower activity in women compared to men, resulting in a longer half-life for substances such as theophylline (metabolised by CYP1A2).

These evaluations do not include changes induced by pregnancy and the possible implications on toxicokinetics and -dynamics.

A long history of reports on increased sensitivity for chemoperception and sensory irritating substances in women is available (Ohla and Lundström 2013). Sex-related differences in susceptibility to sensory irritation, based on self-reported eye irritation ratings, were also reported by Sucker et al. (2019).

3.1.5 Impaired health

In their review Doty et al. (2004) cite several studies which conclude that atopic individuals have a higher than average susceptibility to irritants. Similarly, Shusterman et al. (2003) found that subjects with allergic rhinitis showed a more severe congestive response than normal individuals after exposure to chlorine.

Johansson et al. investigated differences in susceptibility between healthy and asthmatic subjects after inhalation exposure to airborne contaminants (Johansson et al. 2016). The authors evaluated human studies from the literature, in each of which both groups were tested under the same conditions. For 11 substances or mixtures a relevant difference in susceptibility was found, which required assessment factors of >1 to 3 (5 for one study with sulphur dioxide). For four substances no difference in responsiveness was found, for 15 substances the data were not adequate. A detailed comparison of human studies for intensively investigated substances confirmed a higher responsiveness (i.e. lower LOAECs) of individuals with asthma compared to healthy subjects, with quantitative differences in the same range for sulphur dioxide and sulphuric acid but could not identify quantitative differences for ozone and nitrogen dioxide.

3.2 Previous quantifications of variability

3.2.1 Evaluation of toxicokinetic data

Silverman et al. evaluated toxicokinetic data (AUC, Cmax) from clinical Phase I studies from six pharmaceutical active substances to quantify observed inter-individual variability (Silverman et al. 1999). Substance-specific assessment factors for the toxicokinetic part of inter-individual variability were calculated either as the ratio between the 95th percentile of the population and the 50th percentile in case of unimodal distributions (resulting in a factor covering 95% of the population) or, in case of bimodal distributions, as the ratio between the 95th percentile of the sensitive group and the 50th percentile of the general population (covering 95% of the sensitive subgroup).
Two out of six substances yielded factors higher than the WHO standard sub-factor of 3.2: the toxicokinetic factor for amiloride, based on a bimodal distribution, was 5.89. The factor for enalapril was derived from a unimodal distribution and was 3.55. Factors obtained for the other 4 substances ranged from 2.09 to 2.95.

Streeter and Faria extracted human toxicokinetic data (Cmax, AUC, clearance) for 206 pharmaceutical compounds from the literature (Streeter and Faria 2017). Data were obtained for healthy adults and for elderly individuals (not exactly defined). Data were assumed to be normally distributed and for each compound and group (adults, elderly) a factor was calculated as (mean + 2 SD)/mean, intended to cover 95% of the population. However, as distributions for this kind of data are typically skewed to the right, this calculation most likely underestimates the factor and the coverage of the population. All factors calculated for the separate groups of adults and elderly persons were below the subfactor for toxicokinetic variability of 3.2. Therefore, the authors recommended to use a factor of 10 for inter-individual variability of workers to cover

both TK and TD aspects. However, when the differences observed between the two groups were considered in bimodal distributions TK factors were substantially higher (approx. 30% >3.2).

3.2.2 Metabolism-pathway-specific assessment factors (Renwick and Dorne)

In an attempt to develop uncertainty factors specific for certain metabolism pathways, Renwick and Dorne evaluated (mostly clinical phase 1) human studies, published between 1966 and 2003 (Dorne et al. 2001a, b, 2004, 2005; Dorne et al. 2002; Renwick et al. 2001; Renwick and Lazarus 1998). They examined the variability in kinetic parameters (AUC, metabolic and total clearance, Cmax) of pharmaceuticals in healthy adult individuals. Exposure was predominantly oral, but for some substances with variable absorption intravenous data were used. For data analysis the authors assumed that all data are log-normally distributed, and variability was expressed as the coefficient of variation on the logarithmic scale (mean of all data). Two aspects were evaluated

1. Variability within the adult healthy population
2. Differences in toxicokinetic parameters between population groups with polymorph expressed xenobiotica-metabolising enzymes.

Further, differences between different age groups (adults, neonates, children, elder people aged >70) were examined (data not discussed here).

With regard to 1. the variability was expressed as the coefficient of variation on the logarithmic scale (mean of all data). For 2., the ratio of geometric means between extensive and poor metabolisers was calculated. For both, the intraspecies factor required for covering 95 or 99% of the population was calculated, based on the mean variability observed over a range of substances (see **Table 3-1**). To ease comparison with other evaluations (reported below) we added log GSD values calculated from CV_{ln} according to equations given in Dorne et al. (2002).

Log GSD is an important parameter to describe variability within a right-sided dataset and is used, for example, by Hattis and colleagues (Hattis et al. 2002; Hattis and Lynch 2007) and in the IPCS report on uncertainty in hazard characterisation (WHO 2014). It is explained in more detail in Annex 2. There, it is also explained how log GSD can be “translated” into factors intended to cover a certain proportion of the population under study.

Table 3-1 Results on pathway-related factors derived by Dorne et al. (values from Dorne et al. (2005))

Pathway	**CV_{ln}**	**Ratio GM**	**Factor to cover 95th perc.**	**Factor to cover 99th perc.**	**Log GSD calculated from CV_{ln}**
Phase I: Monomorphic pathways (CYP1A2, 2A6, 2E1, 3A4, ADH, Hydrolysis	24 - 46%	-	1.5 - 2.1	1.8 – 2.7	0.10 – 0.16
Phase I: Polymorphic pathways (CYP2C9, 2C19, 2D6)	12 – 66%	1.1 - 31	1.3 - 45	1.5 - 52	
Phase II: Monomorphic pathways (glucuronidation, glycine conjugation, sulphathion)	21 – 29%	-	1.4 – 1.5	1.6 – 2.0	0.09 – 0.12
Phase II: Polymorphic pathways (NAT)	22 – 32%	3.1	1.7 – 4.4	2.1 – 5.2	
Renal excretion	21%	-	1.4	1.6	0.09

The authors concluded that the partial factor of 3.16 used by WHO for the toxicokinetic part of human variability is suitable for substances metabolised by enzymes without known or quantifiable differences due to polymorph expression. But differences in internal body burdens due to polymorphisms exceeded in many cases a factor of 5. The highest factors (up to 52) were calculated for carriers of CYP2C19 with poor metabolic activity. For polymorphisms in CYP2D6 a factor of 26 is required to cover poor metabolisers at the 99th population percentile. Note that these factors were calculated based on geometric means of groups (effective versus poor metabolisers) and do not yet include variability within these groups.

Summarising the results, Dorne (2007) concluded that the factor of 3.2 would not cover substances metabolised via polymorphic isoenzymes. The factor would also not be sufficient to include neonates and the elderly for most elimination routes.

Recently, the data were used in a Bayesian model to predict distributions for route-specific inter-individual differences in toxicokinetics (Wiecek et al. 2019). For 8 exemplary substances metabolised via CYP2D6, uncertainty factors (95th percentile) of 3.1 to 12.2 were derived.

3.2.3 The Hattis database

Hattis and colleagues from the Clark University in Worcester, Massachusetts, USA, were the first to develop an extensive database of inter-individual variability in toxicokinetics and -dynamics (Hattis 1996a, b; Hattis et al. 2002; Hattis et al. 1999a; Hattis et al. 1999b; Hattis and Lynch 2007; Hattis and Silver 1994). The database, openly accessible via the internet[13], was used to develop distributions for inter-individual variability for a probabilistic assessment framework (Hattis and Lynch 2007). A consolidated and updated version, contributed by the authors (and based on the data from Hattis and Lynch), was used as input for the IPCS efforts to create an easy-to-use probabilistic assessment tool (APROBA, Approximated Probabilistic Assessment Tool, see Report "Probabilistic Hazard Assessment") (WHO 2014). Two sets of data, derived from human studies, were used to develop a distribution for differences in toxicokinetics and in toxicodynamics, respectively.

Characterisation of the TK (toxicokinetic) dataset:

For the assessment of toxicokinetic data human studies were evaluated which provide AUC (area under the curve) and Cmax (maximum plasma concentration) values and their variability. The studies comprise Clinical Phase I studies of pharmaceuticals with oral exposure.

Table 3-2 Evaluation of toxicokinetic data from the Hattis database as described in WHO (2014)

Parameter	**N (# sub-stances)**	**Age**	**GM of log GSD_{TK}**	**P95/P50***	**Remark**
AUC	31	Adults	0.161	2.42	
AUC	6	Children <12	0.204	Not given	
Combined AUC data	37	Adults and children	0.167	2.43	Used for distribution in WHO (2014)
Cmax	29	Various ages, includes 5 datasets with <12 years	0.155	2.90	
Combined dataset (AUC and Cmax data)			0.162	2.62	

* Ratio to characterise uncertainty in log GSD, not to confuse with percentiles of population

[13] Hattis database: http://www2.clarku.edu/faculty/dhattis

Characterisation of the TD (toxicodynamics) dataset:

Again, the data as compiled by Hattis and Lynch (2007) were used in the IPCS report (WHO 2014). The initial discussion also included the data on effects in the respiratory tract, showing high variability (see discussion in chapter 3.2.4.2) of (partly) immunologically mediated effects. But these data were not used to derive distributions for toxicodynamic variability. No details on how the logGSD values were derived, is included in Hattis and Lynch (2007) or Hattis et al. (2002); the calculation sheets are currently not accessible at http://www2.clarku.edu/faculty/dhattis. Therefore, the calculation of the values cannot be followed. Substances evaluated are mostly pharmaceuticals applied via the oral route. The following table summarises the data used in the ICPS report (WHO 2014).

Table 3-3 Evaluation of toxicodynamic data from the Hattis database as described in WHO (2014) (table adapted from Table A4.4 in WHO (2014))

Parameter	**N (# sub-stances)**	**GM of log GSD_{TD}**	**P95/P50***	**Remark**
Internal concentration producing specific non-immune related physiological parameter changes	18	0.195	2.76	
Internal concentration producing specific non-immune related quantal responses	16	0.256	2.89	
Combined dataset	34	0.221	2.85	Used for distribution in WHO (2014)
Non-immune related quantal responses in relation to external exposure	10	0.242	4.27	

* Ratio to characterise uncertainty in log GSD, not to confuse with percentiles of population

Note that only data based on internal concentrations were used in WHO (2014), to separate toxicokinetic from –dynamic variability.

According to the evaluations by Hattis et al., children appear to show slightly higher variability with regard to toxicokinetic parameters compared to adults, and the variability with regard to toxicodynamics is higher than for toxicokinetics, as shown by the higher log GSD_{TD} values.

Like Hattis, log GSD is used as a measure of variability in compiling our database in this project for toxicokinetic data. This was described in more detail in chapter 2.1.

In a previous research project for BAuA on probabilistic methods for deriving OELs the Hattis database was used to derive distributions for intraspecies extrapolation (Schneider et al. 2005; Schneider et al. 2006). Using all toxicodynamic data from the Hattis database including the data on airway responsiveness (Hattis et al. 2001) (see chapter 3.2.4.2) distributions reflecting high inter-individual variability were obtained, with factors of

- 19.8 to cover 90% of the population
- 43.8 to cover 95% of the population
- 193.4 to cover 99% of the population

with a 95% probability. As discussed above, these data for toxicodynamic endpoints are not included in the current evaluation.

3.2.4 Local effects in the respiratory tract effects

3.2.4.1 Particle deposition and clearance

Limited information is available on inter-individual differences in deposition and clearance of particles. With lung deposition models, inclusion of information about inter-individual differences in lung anatomy resulted in a threefold difference in airway deposition fraction estimates (Asgharian et al. 2001; Kuempel et al. 2015). Similar, (Löndahl et al. 2014), when reviewing available information on deposition of nanoparticles in the respiratory tract, conclude that there is relevant inter-individual variation in particle deposition, which the authors relate to variability in lung morphology, breathing patterns and other parameters.

Rissler et al. (2017) investigated inter-individual variability in the deposition of airborne particles in a study with 67 healthy volunteers, seven of them between 7 and 12 years old, and 60 adult individuals. The authors measured deposition rates for particles sized 10 to 3500 nm and found substantial differences for all particle sizes. For deposition fractions between 0.4 and 0.8, standard deviations were around 0.1. This means, for example, that at a deposition fraction of 0.5 a range from 0.3 to 0.7 would be required to cover 95% of individuals. Variables related to lung-intrinsic properties (e.g. anatomical airway dead space) and breathing patterns (e.g. time of a breath cycle) best explained the variability. Deposition and clearance may also be altered by pathological states. (Zhang et al. 2018) found that particle kinetics and deposition was altered when modelled considering the anatomical conditions of COPD.

Kuempel et al. (2015) reported that lower clearance rates were observed in retired coal miners compared to healthy adults without occupational dust exposure. Clearance might also be compromised in individuals with chronic obstructive pulmonary disease (COPD). ICRP (International Commission on Radiological Protection) recommended to reduce the clearance rate by a factor of two in such cases (Kuempel et al. 2015).

Gregoratto et al. (2010) evaluated data on clearance in humans from three cohorts of volunteers or workers, who inhaled insoluble, long-living particles:

- ^{195}Au-labelled Teflon particles, inhaled by 10 volunteers
- insoluble particles containing ^{60}Co, accidentally inhaled by workers

- plutonium oxides, accidentally inhaled by workers at the Rocky Flats Plant (RFP).

Large inter-individual differences in clearance were observed in these groups. For the transport rate m_T to the tracheobronchiolar region and the transport rate m_I to the interstitium the authors calculated ranges +/- SD (covering the central 68%) of 0.4 - 4 × 10^{-3} d^{-1} and 0.2 - 4 × 10^{-3} d^{-1}, respectively. This indicates differences in clearance rates of one order of magnitude and more.

3.2.4.2 Health effects in the respiratory tract

Hattis database

Hattis et al. discussed the variability of parameters with influence on the kinetics and toxicity of airborne toxicants (Hattis et al. 2001). As explained above, these authors used log GSD for characterising variability (see Annex 2).

.

Hattis et al. also used log GSD to describe variability of various key physiological parameters in the respiratory tract:

- Breathing activity (derived from activity pattern studies): log GSD = 0.12
- Alveolar deposition of particles (derived from ICRP model application): log GSD = 0.3
- Clearance (based on limited human data on short-term clearance: log GSD = 0.21 (healthy subjects) to 0.34 (impaired health, due to asthma, bronchitis, and other obstructive lung conditions).

These log GSD reflect variabilities similar to those of toxicokinetic parameters for systemically available substances.

Further, the authors evaluated human data on effects in the respiratory tract. These are mainly investigations from airway response provocation tests with methacholine or airway allergens (see e.g. Hanania et al. 1998). Comparison of concentrations resulting in similar effects (e.g. a 20% decrease in FEV_1 – forced expiratory volume in one second) in individuals were compared and revealed large inter-individual differences for these immunology-related effects. For example, Bakke et al. (1991) reported that in the Norwegian population <10% had PC20 values (metacholine concentration inducing a 20% decrease in the forced expiratory volume at 1 second (FEV_1) of <2 mg/mL, whereas >80% had PC20 values of >32 mg/mL (i.e. differences > factor 16). These data resulted in log GSD values of 0.469 (geometric mean for “continuous inhalation parameter change, e.g. FEV1 change)” and 0.550 (for “quantal responses to inhalation, e.g. wheeze, throat irritation)”.

BAuA Research Project 2004

In 2004 FoBiG carried out a project on behalf of BAuA with the aim to gather and evaluate data on intraspecies variability of substances acting predominantly locally in the respiratory tract (FoBiG 2004)**.** Only five substances which show adverse effects in the upper respiratory tract, with enough human data to conclude quantitatively on inter-individual differences were found. For the tracheobronchiolar tract and the lower respiratory tract it was one substance each, only. Overall, a broad range of ratios for

the difference between average susceptible and susceptible individuals (effective concentration average individual/effective concentration susceptible individual) between 2 and 25 was observed. Datasets complying with our acceptance criteria (see 2.1) were included in our database in chapter 2.2.2.

3.2.4.3 Sensory irritation

Brüning et al. (2014) proposed a scheme for deriving OELs based on sensory irritation as observed in human studies and animal experiments. With regard to inter-individual susceptibility for sensory irritation the authors discussed several studies investigating potential differences in susceptibility between groups of chemosensitive or allergic individuals and groups of normal healthy subjects and concluded "*that an intraspecies default factor is not necessary if OELs are derived from human sensory NOAECs since it is based on a controlled human exposure study assessing especially sensitive and objectively verifiable effects*". Sex-related differences or the observed variability within the groups were not discussed. The authors propose to consider inter-individual variability with the lower limit of benchmark dose used as point of departure.

Nielsen and Wolkoff (2017) evaluated studies on sensory irritants for differences between normal individuals and susceptible groups such as atopic subjects or asthmatics. They did not find evidence for pronounced differences but suggest to use a factor of two to account for other factors such as age and sex. Claeson und Lind (2016) found slight differences between "responders" and "non-responders" with regard to perception and eye blink frequencies in a study with acrolein exposure of volunteers.

In an effort to analyse reasons for inter-individual differences in chemosensory perception Pacharra et al. (2016) analysed data from a set of human volunteer studies. Sex was a key determining parameter for differences in ratings for annoyance and disgust (emotions related to a predisposition for high anxiety) as well as for pungency and burning sensations, effects related to olfactory-related sensitivity. A higher susceptibility of women was found for both types of observations, and interdependency with other parameters was observed. Also Sucker et al. (2019) found higher eye irritation ratings reported by women exposed to 5 ppm ethyl acrylate for 4 hours compared to men, but a respective analysis was not presented for the objective measurement of eye blink rates. No influence of atopy on subjective irritation reporting or eye blink rates was observed, but high inter-individual variability was observed for the ethyl acrylate induced effects for the latter endpoint. Similarly, Shusterman and Balmes (1997) found lower thresholds for CO_2-detection as a sign for irritating effects for women and smokers.

3.3 PBPK and IVIVE modelling

PBPK models can predict variability by simulating the impact of inter-individual variability of input parameters, such as renal filtration rate (Krauss et al. 2015). Top-down approaches use data on variability, e.g. from phase 1 clinical trials, and try to

explain the observed variability by combining PBPK modelling with Bayesian methods (Bois et al. 2010; Krishnan et al. 2013). Also, the impact on polymorphisms in metabolising enzymes on delivered dose can be quantified by PBPK modelling (Haber et al. 2002).

In a simplistic modelling approach, which did not make use of any empirical data on observed substance-specific data on inter-individual differences, Nong and Krishnan (2007) estimated the variability in steady state blood concentrations of inhaled volatile organic compounds (benzene, chloroform, methyl chloroform, carbon tetrachloride). The authors obtained subfactors for differences in TK between 1.15 and 1.92.

For performing a cancer risk assessment of dichloromethane Jonsson and Johanson (2001) applied a PBPK model based on and extending the one of El-Masri et al. (1999). Differences in the activity of glutathione S-transferase T1 (GSTT1) and the detoxifying mixed-function oxidases (MFO) pathway contributed most to inter-individual differences in internal body burden. The authors concluded that 5% of the Swedish population would require a toxicokinetic subfactor higher than 2.7 to 3.2 and 1% would need factor higher than 4.2 to 7.1.

Gentry et al. (2002) used PBPK modelling combined with Monte-Carlo (to assess the variability in obtained estimates in AUC) analysis to develop chemical-specific adjustment factors for toxicokinetic variability for the two substances parathion and warfarin. Both substances are metabolised via polymorphic expressed enzymes (CYP2C9 in the case of warfarin and paraoxonase for parathion/paraoxon). For warfarin a high impact of the polymorphism was noted: a factor of 11 was calculated to account for the difference between the median (preferred over the mean for the distinctly skewed distribution) and the 95th percentile of the total population) and a factor of 26 was required at the 99th percentile level. For parathion the respective factors were much smaller: 2.4 for 95th perc/median and 3.2 for 99th perc/median. In a separate calculation the 95th percentile of the sensitive subgroup was compared with the median of the “normal” population, which resulted in a factor of 3.4.

In addition to these, several further publications are available, which combine PBPK modelling with Monte-Carlo analysis or similar techniques to describe variability. These models span a broad range regarding amount and type of input data (animal, in vitro, human), complexity, routes and dose ranges modelled, and verification against observed data. They are not discussed in detail here, but the results obtained and presented in **Table 3-4** give an impression of the variability predicted with PBPK models. As with the data analysed by Dorne et al. (2005) polymorphisms of CYP enzymes (e.g. CYP2D6) yielded the largest differences between individuals. In the review by Valcke and Krishnan (2014) further studies are cited, with subfactors for toxicokinetic variability in the general population including children mostly in the range of 1.6 to 4 and few higher factors (up to 28.3) for substrates metabolised by polymorphic CYP enzymes. In their conclusions the authors state that the subfactor of 3.2 is seldom exceeded.

Table 3-4 Substance-specific prediction of inter-individual variability in healthy adults by published PBPK models (factors as reported by authors to cover variability in 95% and 99% of the population, resp.)

Substance	**Main influencing factors**	**95% of population**	**99% of population**	**Reference**
Dichloro-methane	GSTT1/MFO pathway	2.7 – 3.2	4.2 – 7.1	Jonsson and Johanson (2001)
Dichloro-methane	none	1.9 - 2.0		Pelekis et al. (2003)
Warfarin	CYP2C9 polymorphism	11	26	Gentry et al. (2002)
Parathion	Paraoxonase polymorphism	2.4	3.2	Gentry et al. (2002)
Estragole	Hydroxylation by CYP450 enzymes, oxidation	1‘-hydroxy-estragole 1.4* – 2.7**	1‘-hydroxy-estragole 1.6* – 4.0** 1‘-sulfooxy-estragole 5.4	Punt et al. (2010) Punt et al. (2016)
Methyl-eugenol	CYP4501A2 (hydroxylation), CYP4502B6 (epoxidation)		6.4	Al-Subeihi et al. (2015)
Phenol	UDP-Glucurono-syltransferase 1A6		2.0	(Strikwold et al. 2017)
Toluene		1.6 - 1.7 (chronic cons-tant exposure)	-	(Mörk et al. 2014)

Substance	**Main influencing factors**	**95% of population**	**99% of population**	**Reference**
Styrene		1.6 - 1.7 (chronic constant exposure)	-	(Mörk et al. 2014)
Methyl chloride		1.5 - 1.7 (chronic constant exposure)	-	(Mörk et al. 2014)
Acetone		1.5 – 1.8 (chronic, workers, various conditions)		(Mörk and Johanson 2010)
Volatile organic chemicals	-	1.15 – 1.92	-	Nong and Krishnan (2007)
Bisphenol A		2.4 (based on predicted AUC)		Yang et al. (2015)
Chlorpyrifos			3.4 (adults)	Poet et al. (2017)
Oseltamivir (pro-drug of Ro 64-0802)			(based on predicted AUC) Oseltamivir plasma: 1.8 Ro 64-0802 plasma: 1.5 Oseltamivir brain: 2.5 – 3.8 Ro 64-0802 brain: 2.5 - 5	Ito et al. (2017)

*based on variability observed in 14 humans; ** based on 3fold higher variability

Table 3-4 comprises a broad range of substances, approaches and model sophistications. Accordingly, resulting factors are not directly comparable. However, the figures given in the table describe in a cursory way the ranges of inter-individual toxicokinetic differences predicted by PBPK models, which are 1.15 to 11 at the 95% population level and 1.5 to 26 at the 99% population level.

In vitro – in vivo extrapolation (IVIVE) modelling is another approach, which can help to describe inter-individual variability. Wetmore et al. (2014) used hepatic clearance rates measured in vitro for 13 cytochrome P450 and five uridine 5'-diphospho-glucuronysyltransferase isozymes using recombinantly expressed enzymes and used these clearance rates in an IVIVE model which includes known differences in isozyme expression (cytochorome P450 enzymes and UDP glucuronosyl transferases) in various populations. Differences in steady-state blood concentrations between a healthy population and the upper 95th percentile of sensitive populations modelled for 9 chemicals ranged from 3.1 - to 13.1-fold. In most cases pediatric lifestages up to 0.5 years were the most sensitive ones. Therefore, the quantitative outcome of this modelling exercise cannot be used to describe inter-individual variability in adults.

Wambaugh et al. (2019) performed high throughput in vitro measurements of the unbound plasma fraction and hepatic clearance (by human hepatocytes) for more than 400 substances. The data were used to predict in vivo plasma concentrations in humans (C_{ss}) by IVIVE models. For 389 substances both parameters could be determined and C_{ss} predicted. Uncertainty was modelled for individual substances using the in vitro measurements of the replicate measurements of both unbound plasma fraction and hepatic clearance. Further, the authors used biometric data from the US population (NHANES study) to simulate biological variability in toxicokinetics with Bayesian methods. They used the ratio of the population 95th percentile of C_{ss} divided by the median C_{ss} to characterise both uncertainty and variability and the combined distribution:
Median (for 389 chemicals) ratio for uncertainty (ratio 95th perc./median): 2.32
Median (for 389 chemicals) ratio for variability (ratio 95th perc./median): 6.27
Median (for 389 chemicals) ratio for combined uncertainty and variability: 7.13.

3.4 Population-based data on variability

Descriptions of inter-individual variability due to different genotypes can be obtained at the population level, in principle, from investigations with genetically diverse populations of experimental animals (especially mice, with the caveat of potential differences between species), with primary human cells (for which, however, it is difficult to achieve sufficiently high numbers to assess population variability) and with genetically diverse, immortalised human cell lines, e.g. human lymphoblastoid cell lines (Axelrad et al. 2019; Dornbos and LaPres 2018). Lymphoblastoid cell lines are derived from B cells, which were immortalized by infection with the Epstein-Barr Virus. A large pool (from more than 1000 individuals) of human lymphoblastoid cell lines was established in the frame of the “1000 Genomes project”[14].

[14] https://www.internationalgenome.org/

Mortensen and Euling (2013) outlined how they expect new techniques and databases created in the frame of Tox21 (Krewski et al. 2010; Tice et al. 2013) to improve the knowledge on intraspecies variability. Improved mechanistic data (including omics data) on chemical-gene associations explain how specific genes affect genetic susceptibility and data from DNA sequencing projects help to identify variations in genes which contribute to variability in chemical-target interactions in the general population. They cited benzene as an example, for which mechanistic, genetic and epigenetic susceptibility data could explain differences in toxicokinetics (certain combinations of CYP2E1 and NAD(P)H dehydrogenase polymophisms increase the production of toxic metabolites) and toxicodynamics (by genes related to DNA repair and genomic maintenance). The authors also refer to Fry et al. (2008), who by investigating variations in transcription profiles in 24 human lymphoblastoid cell lines, characterised the variability in susceptibility to DNA alkylating agents.

Human lymphoblastoid cell lines were also used by Abdo et al. (2015) for characterising the variation in in vitro cytotoxicity for 179 chemicals. The chemicals are part of the National Toxicology Program's chemical library. The 1086 cell lines tested were from individuals from five continents and nine populations ("1000 Genomes project", Coriell Institute). The chemicals were tested for cytotoxicity (intracellular ATP) at 8 different concentrations over 6 orders of magnitude and an ED10 was determined by curve fitting. Variability for each substance was described by percentiles of the obtained empirical distributions. Further, factors were calculated for each dataset describing the difference between the 1st (or 5th percentile) and the median, reflecting the difference in response of the 1% (or 5%) with the lowest ED10 (highest susceptibility) and the median. These "raw factors" were corrected for sampling variability (variation between replicate measurements), which reduced variability considerably. The median for the distribution of ratio of median/1st percentile was 3.04 (90th CI 1.48 – 10.3). The authors compared their results with the distribution used in the IPCS report (WHO 2014), based on Hattis et al (see chapter 3.2.3). The authors also identified some genes associated with high inter-individual variability, among them several genes coding for membrane-bound solute carrier proteins.

Table 3-5 Statistical characterisation of dose ratios reported by Abdo et al. (2015)

Distribution	**Median**	**5th percentile**	**95th percentile**
Median/5th percentile* (95% of population)	1.95	1.19	4.67
Median/1st percentile (99% of population)	3.04	1.44	10.32

*own recalculation from substance-specific data provided in Supplemental material

This study provides information on variability in the toxic response in immortalised cells from over 1000 individuals representing different populations from five different continents. The 179 chemicals tested were from the NTP chemical library and thus reduce the uncertainty regarding representativeness of chemicals (compared to

datasets mainly consisting of data from pharmaceuticals). Furthermore, the high-throughput application allowed to test replicates and correct results for the experimental uncertainty.
Disadvantages are the limited metabolic capacity of the lymphoblastoid cell lines. Furthermore, the in vitro approach might not be conservative as in vivo additional pathways such as epigenetic changes, inhibition of repair enzymes etc. might be influenced by the substances and lead to higher differences between individuals. Also, only one endpoint, namely internal ATP production, was investigated by Abdo et al. (2015). Dornbos and LaPres (2018) noticed that some substances showed different variability for different in vitro endpoints (ATP production versus caspase 3/7 activity).

With the data as obtained by Abdo et al. (2015) and by using different methods for fitting the data to a log-normal distribution (Chiu et al. 2017) for a selection of 138 substances from the Abdo-dataset obtained a slightly different distribution for the ratio median/1st percentile with a median of 2.48 (90% CI 1.44 – 9.57). These authors advocate to use these population-based in vitro data on variability for risk assessment as part of probabilistic approaches (Chiu and Rusyn 2018).

4 Discussion of suitable distributions for describing inter-individual variability

4.1 Variability in toxicokinetics

Our own data evaluation as well as literature data clearly indicate that there is inter-individual variability in internal doses at a given external exposure. Various lines of evidence exist:

- Substance-specific extrapolation factors (covering the difference between medians and the 95th percentile of the population) in the literature derived based on toxicokinetic data mostly ranged from 2 to 6 (chapter 3.2.1)
- metabolism pathway-specific subfactors derived by Dorne and colleagues were mostly in the range 1.5 to 5, but pathways associated with genetic polymorphisms of CYP2C9, CYP2C19, or CYP2D6 required much higher factors, up to 52 (for covering 99% of the population) (chapter 3.2.2)
- The importance of certain polymorphisms was confirmed by substance-specific modelling of kinetics with PBPK models: for most substances toxicokinetic intraspecies factors of 1.5 to 5 were obtained, but for substances metabolised by certain polymorphic enzymes (mostly Phase I CYP) factors as large as 26 were necessary to cover 99% of the population (chapter 3.3)
- The Hattis database provides a substantial amount of toxicokinetic datasets (chapter 3.2.3); its most recent form was used in the IPCS report (WHO 2014) and contained 37 datasets, a few of them from children for (mainly) oral administration of pharmaceuticals; the log GSD distribution (median: 0.162) derived from that database is discussed below
- The new dataset created in this project (chapter 2.2.1) consists of 68 datasets from adults only, including 33 with oral and 31 datasets with inhalation exposure (median of log GSD for total dataset: 0.146; oral data: 0.167; inhalation data: 0.106).

The Hattis database mainly consists of data from studies with pharmaceuticals and oral exposure. In contrast, we made efforts to find and include to a larger extent industrial chemicals and inhalation exposure data.
As explained in chapter 2.2.1.5 there is a correlation of the two parameters “industrial chemical” and “inhalation exposure”: most inhalation data were indeed obtained from studies with industrial chemicals. This is not surprising as oral administration is more important for pharmaceuticals and easier to control in clinical studies, whereas inhalation exposure is a priority consideration for industrial chemicals.
Significant differences were observed for data from oral administration/ pharmaceuticals versus inhalation exposure/industrial chemicals. Currently we are unable to decide whether the substance type, the route of administration or another (associated) condition is critical for this difference. In the following discussion we stratify according to route of exposure.

The following table compares the distributions of log GSD values from the Hattis database as used in (WHO 2014) with the distributions obtained from our database. The factors listed in the table are extrapolation factors describing the distance between the median of the population and the 95th percentile, calculated from log GSD values and to ease interpretation of log GSD values. As mentioned earlier, the concept of log GSD to describe variability is explained in more detail in Annex 2. There, also the calculation of factors to cover certain percentiles of the population is explained.

Table 4-1 Distributions of log GSD

Source	**Dataset**	**N**	**Median of log GSD**	**Equating to factor covering 95% of population***	**Equating to factor covering 99% of population***	**95th per-centile of log GSD**	**Equating to factor covering 95% of population***	**Equating to factor covering 99% of population***
Hattis database (WHO 2014)	mostly oral, pharmaceuticals	37	0.167	1.88	2.45	0.407	4.67	8.85
Our database	Total dataset	68	0.146	1.74	2.19	0.355	3.84	6.70
Our database	Oral (mostly pharmaceuticals)	33	0.167	1.88	2.45	0.379	4.20	7.62
Our database	Inhalation (mostly industrial chemicals)	31	0.106	1.49	1.76	0.252	2.60	3.86

* Calculated according to algorithms as described in IPCS report (WHO 2014); see also the example calculation in Annex 2

It is obvious that the dataset for the oral route from our evaluation is very similar to the Hattis distribution, which itself is mainly based on oral data from pharmaceuticals. **Table 4-2** provides an overview on all information sources listed further above. The factors in the table describe the distance between the median of the population and the 95th or 99th percentile, resp.

Table 4-2 Overview on human variability data in toxicokinetics

Source	**Type of data/evaluation**	**Factor to cover 95% of population**	**Factor to cover 99% of population**
Silverman et al. (1999)	Toxicokinetic data for 6 pharmaceuticals	2.09 - 5.89	
Data compilation by Dorne et al.	Pathway-related factors	1.3 – 45	1.5 - 52
PBPK models	Range of factors obtained from models for 14 substances or substance groups (**Table 3-4**)	1.15 - 11	1.6 – 26
IVIVE modelling (Wambaugh et al. 2019)	Median of ratio 95th perc/median for 389 chemicals	6.27	
Hattis database, toxicokinetic distribution as proposed in IPCS report (WHO 2014)	- Median - 95% probability	1.88* 4.67*	2.45* 8.85*

Source	Type of data/evaluation	Factor to cover 95% of population	Factor to cover 99% of population
This evaluation	Medians:		
	Full dataset	1.74*	2.19*
	Oral data	1.88*	2.45*
	Inhalation data	1.49*	1.76*

* Calculated according to algorithms as described in IPCS report (WHO 2014); see also the example calculation in Annex 2

The reported data show a rather high agreement. However, very high variability is observed for some chemicals metabolised by polymorphically expressed enzymes, such as CYP2D6. The percentage of substances showing such high variability in the chemical universe is unknown. Also, it is difficult to assess whether they are sufficiently represented in our evaluation.

To our knowledge, no consistent dataset existed so far for describing toxicokinetic variability after inhalation exposure. The observation that inter-individual differences might be lower after inhalation than after oral exposure was not reported in the literature before. We recommend to confirm the observed difference between oral and inhalation exposure by expanding the database on inhalation studies.

The 95th percentile of the log GSD distribution of 0.355 corresponds to an assessment factor of 3.84 to cover the distance between the population median and the 95th percentile of the exposed. For the distance between the median and the 99th percentile of the population the factor is 6.7. In comparison with the data reported by Dorne et al. (2005) higher values above 10 resulting from polymorphisms seem not to be represented in sufficient number to influence the 95th percentile of the distribution. Such extreme cases might be underestimated by the distribution from our evaluation. Furthermore, for the evaluation of datasets representing combined toxicokinetic and –dynamic variability (chapter 4.2) inhalation data showed higher variability than oral data.

Considering these uncertainties we recommend

- to increase the toxicokinetic database with inhalation exposure, with a focus to identify substances metabolised by polymorphic metabolising enzymes
- to use our combined (all routes) dataset of 68 substances to characterize intraspecies variability.

4.2 Variability in toxicodynamics

Less data is available to characterise toxicodynamic reasons for differences in susceptibility. Again, the Hattis database was used in the IPCS report on uncertainty to derive a distribution (WHO 2014). However, few details on how the log GSD values were calculated are available and the original calculation files are not accessible currently. Therefore, there remain some uncertainties.

The evaluation performed in our project identified 24 datasets, but the quantification of differences in susceptibility is also considered uncertain, due to reasons discussed above (chapter 2.2.2.3). A broad range of ratios was obtained. This broad range is mainly produced by industrial chemicals after inhalation, which caused irritating effects in the respiratory tract (11 out of 12 datasets with inhalation exposure).

An interesting new dataset was described by Abdo et al. (2015). The authors provided information on variability in the toxic response to 179 chemicals in immortalised cells from over 1000 individuals representing different populations from five different continents. Advantages and disadvantages were discussed above (chapter 3.4).

Some differences between sexes are reported for sensory irritating effects, although it is assumed that inter-individual differences are low for these types of effects. A more detailed analyses on this subject is recommended.

The table below provides an overview on these data. The factors listed describe the distance between the median of the population and the 95^{th} or 99^{th} percentile, resp. In addition, in the table log GSD values from the Hattis evaluation are translated into such extrapolation factors for easy comparison with data reported by others. As mentioned earlier, the concept of log GSD to describe variability is explained in more detail in Annex 2. There, also the calculation of factors to cover certain percentiles of the population is explained.

Table 4-3 Overview on human variability data in toxicodynamics

Source	**Type of data/ evaluation**	**Factor to cover 95% of popula-tion**	**Factor to cover 99% of popula-tion**	**Observed variability ranges (max/min)**
Hattis database, toxicodynamic distribution as proposed in IPCS report (WHO 2014)	- Median - 95% probability	2.31* 10.91*	3.27* 29.37*	
Population-based toxicodynamic variability, based on in vitro high throughput data evaluation by (Abdo et al. 2015)	- Median - 95% probability	1.95** 4.67**	3.04** 10.32**	
This evaluation	Full dataset (N=24)			3 to 201 (large uncertainty)
Sensory irritation studies	Higher susceptibility of women and smokers? Higher susceptibility of atopic or asthmatic individuals?			Yes, quanti-fication unsure No clear indication

Source	**Type of data/ evaluation**	**Factor to cover 95% of popula-tion**	**Factor to cover 99% of popula-tion**	**Observed variability ranges (max/min)**
Irritating effects in the respiratory tract This evaluation Published data on - Particle deposition - Clearance	Dataset local effects (N=11)			4 to 201 (large uncertainty) 2 – 3 (large uncertainty) 2 – 10 (large uncertainty)

* Calculated according to algorithms as described in IPCS report (WHO 2014) and in Annex 2

** Calculated from data reported in Supplementary Material of Abdo et al. (2015)

The first two lines of the table show that similar, but slightly lower factors result from the Abdo et al. (2015) data. But in general, the Hattis data and the data from Abdo et al. (2015) are in good agreement. Considering the broad coverage of the Abdo data in terms of individuals and substances, it is proposed to use this dataset for developing a distribution for toxicodynamic differences in susceptibility.

Fewer information exists on variability for local effects in the respiratory tract after inhalation. Our evaluation indicates that large differences in susceptibility also exist for irritating effects in the respiratory tract. However, for reasons discussed above, quantification is associated with high uncertainties. But it can be concluded from these data that variability is at least as high as observed in vitro by Abdo et al. (2015).

References

Abdo, N., Xia, M., Brown, C.C., Kosyk, O., Huang, R., Sakamuru, S., Zhou, Y.H., Jack, J.R., Gallins, P., Xia, K., Li, Y., Chiu, W.A., Motsinger-Reif, A.A., Austin, C.P., Tice, R.R., Rusyn, I., Wright, F.A. (2015). Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 genomes high-throughput screening study. Environmental Health Perspectives 123, 458-466.

Al-Subeihi, A.A.A., Alhusainy, W., Kiwamoto, R., Spenkelink, B., van Bladeren, P.J., Rietjens, I.M.C.M., Punt, A. (2015). Evaluation of the interindividual human variation in bioactivation of methyleugenol using physiologically based kinetic modeling and Monte Carlo simulations. Toxicology and Applied Pharmacology 283, 117-126.

Asgharian, B., Hofmann, W., Bergmann, R. (2001). Particle deposition in a multiple-path model of the human lung. Aerosol Science and Technology 34, 332-339.

Axelrad, D.A., Setzer, R.W., Bateson, T.F., DeVito, M., Dzubow, R.C., Fitzpatrick, J.W., Frame, A.M., Hogan, K.A., Houck, K., Stewart, M. (2019). Methods for evaluating variability in human health dose-response characterization. Human and Ecological Risk Assessment 25, 1-24.

Bakke, P.S., Baste, V., Gulsvik, A. (1991). Bronchial responsiveness in a Norwegian community. American Review of Respiratory Disease 143, 317-322.

Bhat, V.S., Meek, M.E., Valcke, M., English, C., Boobis, A., Brown, R. (2017). Evolution of chemical-specific adjustment factors (CSAF) based on recent international experience; increasing utility and facilitating regulatory acceptance. Critical Reviews in Toxicology 47, 729-749.

Bois, F.Y., Jamei, M., Clewell, H.J. (2010). PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. Toxicology 278, 256-267.

Brüning, T., Bartsch, R., Bolt, H.M., Desel, H., Drexler, H., Gundert-Remy, U., Hartwig, A., Jäckh, R., Leibold, E., Pallapies, D., Rettenmeier, A.W., Schlüter, G., Stropp, G., Sucker, K., Triebig, G., Westphal, G., van Thriel, C. (2014). Sensory irritation as a basis for setting occupational exposure limits. Archives of Toxicology 88, 1855-1879.

Chen, M.L., Lee, S.C., Ng, M.J., Schuirmann, D.J., Lesko, L.J., Williams, R.L. (2000). Pharmacokinetic analysis of bioequivalence trials: implications for sex-related issues in clinical pharmacology and biopharmaceutics. Clinical Pharmacology and Therapeutics 68, 510-521.

Chiu, W.A., Rusyn, I. (2018). Advancing chemical risk assessment decision-making with population variability data: challenges and opportunities. Mammalian Genome 29, 182-189.

Chiu, W.A., Wright, F.A., Rusyn, I. (2017). A tiered, Bayesian approach to estimating of population variability for regulatory decision-making. ALTEX - Alternativen zu Tierexperimenten 34, 377-388.

Chriguer, R.S., Elias, L.L., da Silva, I.M., Jr., Vieira, J.G., Moreira, A.C., de Castro, M. (2005). Glucocorticoid sensitivity in young healthy individuals: *in vitro* and *in vivo* studies. Journal of Clinical Endocrinology and Metabolism 90, 5978-5984.

Claeson, A.-S., Lind, N. (2016). Human exposure to acrolein: Time-dependence and individual variation in eye irritation. Environmental Toxicology and Pharmacology 45, 20-27.

Clewell, H.J., Teeguarden, J., McDonald, T., Sarangapani, R., Lawrence, G., Covington, T., Gentry, R., Shipp, A. (2002). Review and evaluation of the potential impact of age- and gender-specific pharmacokinetic differences on tissue dosimetry. Critical Reviews in Toxicology 32, 329-389.

Dornbos, P., LaPres, J.J. (2018). Incorporating population-level genetic variability within laboratory models in toxicology: From the individual to the population. Toxicology 395, 1-8.

Dorne, J.L.C.M. (2007). Human variability in hepatic and renal elimination: implications for risk assessment. Journal of Applied Toxicology 27, 411-420.

Dorne, J.L.C.M., Walton, K., Renwick, A.G. (2001a). Human variability in glucuronidation in relation to uncertainty factors for risk assessment. Food and Chemical Toxicology 39, 1153-1173.

Dorne, J.L.C.M., Walton, K., Renwick, A.G. (2001b). Uncertainty factors for chemical risk assessment: Human variability in the in vivo pharmacokinetics of CYP1A2 substrates. Food and Chemical Toxicology 39, 681-696.

Dorne, J.L.C.M., Walton, K., Renwick, A.G. (2004). Human variability in the renal elimination of foreign compounds and renal excretion-related uncertainty factors for risk assessment. Food and Chemical Toxicology 42, 261-274.

Dorne, J.L.C.M., Walton, K., Renwick, A.G. (2005). Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review. Food and Chemical Toxicology 43, 203-216.

Dorne, J.L.C.M., Walton, K., Slob, W., Renwick, A.G. (2002). Human variability in polymorphic CYP2D6 metabolism: is the kinetic default uncertainty factor adequate? Food and Chemical Toxicology 40, 1633-1656.

Doty, R.L., Cometto-Muniz, J.E., Jalowayski, A.A., Dalton, P., Kendal-Reed, M., Hodgson, M. (2004). Assessment of upper respiratory tract and ocular irritative effects of volatile chemicals in humans. Critical Reviews in Toxicology 34, 85-142.

ECHA, European Chemicals Agency. (2012). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1, November 2012, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

El-Masri, H.A., Bell, D., Portier, C.J. (1999). Effects of glutathione transferase theta polymorphism on the risk estimates of dichloromethane to humans. Toxicology and Applied Pharmacology 158, 221-230.

Falk-Filipsson, A., Hanberg, A., Victorin, K., Warholm, M., Wallén, M. (2007). Assessment factors - applications in health risk assessment of chemicals. Environmental Research 104, 108-127.

FoBiG, Forschungs- und Beratungsinstitut Gefahrstoffe GmbH. (2004). Intraspeziesvariabilität bei lokal wirksamen Substanzen – Ansätze zur Quantifizierung eines Defaultfaktors, Im Auftrag der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund.

Fry, R.C., Svensson, J.P., Valiathan, C., Wang, E., Hogan, B.J., Bhattacharya, S., Bugni, J.M., Whittaker, C.A., Samson, L.D. (2008). Genomic predictors of interindividual differences in response to DNA damaging agents. Genes & Development 22, 2621-2626.

Garcia, G.J.M., Schroeter, J.D., Segal, R.A., Stanek, J., Foureman, G.L., Kimbell, J.S. (2009). Dosimetry of nasal uptake of water-soluble and reactive gases: A first study of interhuman variability. Inhalation Toxicology 21, 607-618.

Gentry, P.R., Hack, C.E., Haber, L., Maier, A., Clewell, H.J. (2002). An approach for the quantitative consideration of genetic polymorphism data in chemical risk assessment: examples with warfarin and parathion. Toxicological Sciences 70, 120-139.

Gochfeld, M. (2007). Framework for gender differences in human and animal toxicology. Environmental Research 104, 4-21.

Gochfeld, M. (2016). Sex differences in human and animal toxicology: toxicokinetics. Toxicologic Pathology 45, 172-189.

Gomez, A., Ingelman-Sundberg, M. (2009). Pharmacoepigenetics: its role in interindividual differences in drug response. Clinical Pharmacology and Therapeutics 85, 426-430.

Gregoratto, D., Bailey, M.R., Marsh, J.W. (2010). Modelling particle retention in the alveolar–interstitial region of the human lungs. Journal of Radiological Protection 30, 491-512.

Haber, L.T., Maier, A., Gentry, P.R., Clewell, H.J., Dourson, M.L. (2002). Genetic polymorphisms in assessing interindividual variability in delivered dose. Regulatory Toxicology and Pharmacology 35, 177-197.

Hanania, N.A., Tarlo, S.M., Silverman, F., Urch, B., Senathirajah, N., Zamel, N., Corey, P. (1998). Effect of exposure to low levels of ozone on the response to inhaled allergen in allergic asthmatic patients. Chest 114, 752-756.

Hattis, D. (1996a). Human interindividual variability in susceptibility to toxic effects: from annoying detail to a central determinant of risk. Toxicology 111, 5-14.

Hattis, D. (1996b). Variability in susceptibility - how big, how often, for what responses to what agents? Environmental Toxicology and Pharmacology 2, 135-145.

Hattis, D., Baird, S., Goble, R. (2002). A Straw Man Proposal for a Quantitative Definition of the RfD, in Final Technical Report, U.S. Environmental Protection Agency STAR grant # R825360, Human Variability in Parameters Potentially Related to Susceptibility for Noncancer Risks. Full version available at http://www2.clarku.edu/faculty/dhattis ; shortened version in Drug and Chemical Toxicology 2002; 25:403-436.

Hattis, D., Banati, P., Goble, R. (1999a). Distributions of individual susceptibility among humans for toxic effects. Annals of the New York Academy of Sciences 895, 286-316.

Hattis, D., Banati, P., Goble, R., Burmaster, D.E. (1999b). Human interindividual variability in parameters related to health risks. Risk Analysis 19, 711-726.

Hattis, D., Lynch, M.K. (2007). Empirically Observed Distributions of Pharmacokinetic and Pharmacodynamic Variability in Humans Implications for the Derivation of Single Point Component Uncertainty Factors Providing Equivalent Protection as Existing RfDs. In: J.C. Lipscomb, E.V. Ohanian (Eds), Toxicokinetics in risk assessment, pp. 69-93.

Hattis, D., Russ, A., Goble, R., Banati, P., Chu, M. (2001). Human interindividual variability in susceptibility to airborne particles. Risk Analysis 21, 585-600.

Hattis, D., Silver, K. (1994). Human interindividual variability - a major source of uncertainty in assessing risks for noncancer health effects. Risk Analysis 14, 421-431.

Hine, C.H., Kodama, J.K., Guzman, R.J., Loquvam, G.S. (1960). The toxicity of allylamines. Archives of Environmental Health 1, 343-352.

Ito, M., Kusuhara, H., Ose, A., Kondo, T., Tanabe, K., Nakayama, H., Horita, S., Fujita, T., Sugiyama, Y. (2017). Pharmacokinetic modeling and Monte Carlo simulation to predict interindividual variability in human exposure to oseltamivir and its active metabolite, Ro 64-0802. AAPS Journal 19, 286-297.

Ivanov, M., Kacevska, M., Ingelman-Sundberg, M. (2012). Epigenomics and interindividual differences in drug response. Clinical Pharmacology and Therapeutics 92, 727-736.

Johansson, M.K., Johanson, G., Öberg, M. (2016). Evaluation of the experimental basis for assessment factors to protect individuals with asthma from health effects during short-term exposure to airborne chemicals. Critical Reviews in Toxicology 46, 241-260.

Jonsson, F., Johanson, G. (2001). A Bayesian analysis of the influence of GSTT1 polymorphism on the cancer risk estimate for dichloromethane. Toxicology and Applied Pharmacology 174, 99-112.

Kozyra, M., Ingelman-Sundberg, M., Lauschke, V.M. (2017). Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. Genetics in Medicine 19, 20-29.

Krauss, M., Tappe, K., Schuppert, A., Kuepfer, L., Goerlitz, L. (2015). Bayesian population physiologically-based pharmacokinetic (PBPK) approach for a physiologically realistic characterization of interindividual variability in clinically relevant populations. PLoS One 10, e0139423.

Krewski, D., Acosta, D., Andersen, M., Anderson, H., Bailar, J.C., Boekelheide, K., Brent, R., Charnley, G., Cheung, V.G., Green, S., Kelsey, K.T., Kerkvliet, N.I., Li, A.A., McCray, L., Meyer, O., Patterson, R.D., Pennie, W., Scala, R.A., Solomon, G.M., Stephens, M., Yager, J., Zeise, L. (2010). Toxicity testing in the 21st century: a vision and a strategy. Journal of Toxicology and Environmental Health. Part B 13, 51-138.

Krishnan, K., McPhail, B., Chiu, W., White, P. (2013). Chapter 4 - Modeling of sensitive subpopulations and interindividual variability in pharmacokinetics for health risk assessments. In: B.A. Fowler (Ed), Computational Toxicology. Methods and Applications for Risk Assessment, San Diego, pp. 45-66.

Kuempel, E.D., Sweeney, L.M., Morris, J.B., Jarabek, A.M. (2015). Advances in inhalation dosimetry models and methods for occupational risk assessment and exposure limit derivation. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S18-40.

Lee, H.H., Ho, R.H. (2017). Interindividual and interethnic variability in drug disposition: polymorphisms in organic anion transporting polypeptide 1B1 (OATP1B1; SLCO1B1). British Journal of Clinical Pharmacology 83, 1176-1184.

Löndahl, J., Möller, W., Pagels, J.H., Kreyling, W.G., Swietlicki, E., Schmid, O. (2014). Measurement techniques for respiratory tract deposition of airborne

nanoparticles: a critical review. Journal of Aerosol Medicine and Pulmonary Drug Delivery 27, 229-254.

Mörk, A.-K., Johanson, G. (2010). Chemical-specific adjustment factors for intraspecies variability of acetone toxicokinetics using a probabilistic approach. Toxicological Sciences 116, 336-348.

Mörk, A.-K., Jonsson, F., Johanson, G. (2014). Adjustment factors for toluene, styrene and methyl chloride by population modeling of toxicokinetic variability. Regulatory Toxicology and Pharmacology 69, 78-90.

Mortensen, H.M., Euling, S.Y. (2013). Integrating mechanistic and polymorphism data to characterize human genetic susceptibility for environmental chemical risk assessment in the 21st century. Toxicology and Applied Pharmacology 271, 395-404.

Naumann, B.D., Dolan, D.G., Sargent, E.V. (2004). Rationale for the chemical-specific adjustment factors used to derive an occupational exposure limit for timolol maleate. Human and Ecological Risk Assessment: An International Journal 10, 99-111.

Neafsey, P., Ginsberg, G., Hattis, D., Johns, D.O., Guyton, K.Z., Sonawane, B. (2009). Genetic polymorphism in CYP2E1: Population distribution of CYP2E1 activity. Journal of Toxicology and Environmental Health. Part B 12, 362-388.

Nebert, D.W. (2005). Inter-individual susceptibility to environmental toxicants - A current assessment. Toxicology and Applied Pharmacology 207, Suppl. 1, 34-42.

Nicolas, J.-M., Espie, P., Molimard, M. (2009). Gender and interindividual variability in pharmacokinetics. Drug Metabolism Reviews 41, 408-421.

Nielsen, G.D., Wolkoff, P. (2017). Evaluation of airborne sensory irritants for setting exposure limits or guidelines: A systematic approach. Regulatory Toxicology and Pharmacology 90, 308-317.

Ning, J., Louisse, J., Spenkelink, B., Wesseling, S., Rietjens, I.M.C.M. (2017). Study on inter-ethnic human differences in bioactivation and detoxification of estragole using physiologically based kinetic modeling. Archives of Toxicology 91, 3093-3108.

Nong, A., Krishnan, K. (2007). Estimation of interindividual pharmacokinetic variability factor for inhaled volatile organic chemicals using a probability-bounds approach. Regulatory Toxicology and Pharmacology 48, 93-101.

Ohla, K., Lundström, J. (2013). Sex differences in chemosensation: sensory or emotional? Frontiers in Human Neuroscience 7.

Pacharra, M., Kleinbeck, S., Schäper, M., Juran, S.A., Hey, K., Blaszkewicz, M., Lehmann, M.L., Golka, K., van Thriel, C. (2016). Interindividual differences in

chemosensory perception: Toward a better understanding of perceptual ratings during chemical exposures. Journal of Toxicology and Environmental Health. Part A 79, 1026-1040.

Pelekis, M., Nicolich, M.J., Gauthier, J.S. (2003). Probabilistic framework for the estimation of the adult and child toxicokinetic intraspecies uncertainty factors. Risk Analysis 23, 1239-1255.

Poet, T.S., Timchalk, C., Bartels, M.J., Smith, J.N., McDougal, R., Juberg, D.R., Price, P.S. (2017). Use of a probabilistic PBPK/PD model to calculate Data Derived Extrapolation Factors for chlorpyrifos. Regulatory Toxicology and Pharmacology 86, 59-73.

Punt, A., Jeurissen, S.M., Boersma, M.G., Delatour, T., Scholz, G., Schilter, B., van Bladeren, P.J., Rietjens, I.M. (2010). Evaluation of human interindividual variation in bioactivation of estragole using physiologically based biokinetic modeling. Toxicological Sciences 113, 337-348.

Punt, A., Paini, A., Spenkelink, A., Scholz, G., Schilter, B., van Bladeren, P.J., Rietjens, I.M.C.M. (2016). Evaluation of interindividual human variation in bioactivation and DNA adduct formation of estragole in liver predicted by physiologically based kinetic/dynamic and Monte Carlo Modeling. Chemical Research in Toxicology 29, 659-668.

Renwick, A.G., Dorne, J.L., Walton, K. (2001). Pathway-related factors: The potential for human data to improve the scientific basis of risk assessment. Human and Ecological Risk Assessment 7, 165-180.

Renwick, A.G., Lazarus, N.R. (1998). Human variability and noncancer risk assessment - An analysis of the default uncertainty factor. Regulatory Toxicology and Pharmacology 27, 3-20.

Rissler, J., Gudmundsson, A., Nicklasson, H., Swietlicki, E., Wollmer, P., Londahl, J. (2017). Deposition efficiency of inhaled particles (15-5000 nm) related to breathing pattern and lung function: an experimental study in healthy children and adults. Particle and Fibre Toxicology 14, 10.

Schneider, K., Hassauer, M., Oltmanns, J., Schuhmacher-Wolz, U., Elmshäuser, E., Mosbach-Schulz, O. (2005). Uncertainty Analysis in Workplace Effect Assessment. Research Report F 1824, F 1825, F 1826, Federal Institute for Occupational Safety and Health Dortmund Germany.

Schneider, K., Schuhmacher-Wolz, U., Hassauer, M., Darschnik, S., Elmshäuser, E., Mosbach-Schulz, O. (2006). A probabilistic effect assessment model for hazardous substances at the workplace. Regulatory Toxicology and Pharmacology 44, 172-181.

Shusterman, D., Balmes, J. (1997). Measurement of nasal irritant sensitivity to pulsed carbon dioxide: a pilot study. Archives of Environmental Health 52, 334-340.

Shusterman, D., Murphy, M.A., Balmes, J. (2003). Influence of age, gender, and allergy status on nasal reactivity to inhaled chlorine. Inhalation Toxicology 15, 1179-1189.

Silverman, K.C., Naumann, B.D., Holder, D.J., Dixit, R., Faria, E.C., Sargent, E.V., Gallo, M.A. (1999). Establishing data-derived adjustment factors from published pharmaceutical clinical trial data. Human and Ecological Risk Assessment 5, 1059-1089.

Streeter, A.J., Faria, E.C. (2017). Analysis of the variability of the pharmacokinetics of multiple drugs in young adult and elderly subjects and its implications for acceptable daily exposures and cleaning validation limits. International Journal of Hygiene and Environmental Health 220, 659-672.

Strikwold, M., Spenkelink, B., Woutersen, R.A., Rietjens, I.M., Punt, A. (2017). Development of a combined *in vitro* physiologically based kinetic (PBK) and Monte Carlo modelling approach to predict interindividual human variation in phenol-induced developmental toxicity. Toxicological Sciences 157, 365-376.

Sucker, K., Hoffmeyer, F., Monse, C., Jettkant, B., Berresheim, H., Rosenkranz, N., Raulf, M., Bunger, J., Bruning, T. (2019). Ethyl acrylate: influence of sex or atopy on perceptual ratings and eye blink frequency. Archives of Toxicology 93, 2913-2926.

Thummel, K.E., Lin, Y.S. (2014). Sources of interindividual variability. In: Enzyme Kinetics in Drug Metabolism. Methods in Molecular Biology (Methods and Protocols). Vol. 1113 (Ed), Nagar, S.; Argikar, U.; Tweedie, D., Humana Press, Totowa, NJ, pp. 363-415.

Tice, R.R., Austin, C.P., Kavlock, R.J., Bucher, J.R. (2013). Improving the human hazard characterization of chemicals: a Tox21 update. Environmental Health Perspectives 121, 756-765.

Tracy, T.S., Chaudhry, A.S., Prasad, B., Thummel, K.E., Schuetz, E.G., Zhong, X.B., Tien, Y.C., Jeong, H., Pan, X., Shireman, L.M., Tay-Sontheimer, J., Lin, Y.S. (2016). Interindividual variability in cytochrome P450-mediated drug metabolism. Drug Metabolism and Disposition 44, 343-351.

US EPA, Environmental Protection Agency. (2014). Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. EPA/100/R-14/002F, U.S. Environmental Protection Agency, Risk Assessment Forum, Office of the Science Advisor, Washington, DC. https://www.epa.gov/sites/production/files/2015-01/documents/ddef-final.pdf.

Valcke, M., Krishnan, K. (2014). Characterization of the human kinetic adjustment factor for the health risk assessment of environmental contaminants. Journal of Applied Toxicology 34, 227-240.

Wambaugh, J.F., Wetmore, B.A., Ring, C.L., Nicolas, C.I., Pearce, R.G., Honda, G.S., Dinallo, R., Angus, D.S., Gilbert, J., Sierra, T., Badrinarayanan, A., Snodgrass, B., Brockman, A., Strock, C., Setzer, R.W., Thomas, R.S. (2019). Assessing toxicokinetic uncertainty and variability in risk prioritization. Toxicological Sciences 172, 235-251.

Wenker, M.A., Kezic, S., Monster, A.C., de Wolff, F.A. (2001). Metabolic capacity and interindividual variation in toxicokinetics of styrene in volunteers. Human and Experimental Toxicology 20, 221-228.

Wetmore, B.A., Allen, B., Clewell, H.J., Parker, T., Wambaugh, J.F., Almond, L.M., Sochaski, M.A., Thomas, R.S. (2014). Incorporating population variability and susceptible subpopulations into dosimetry for high-throughput toxicity testing. Toxicological Sciences 142, 210-224.

WHO, World Health Organization. (1994). Environmental Health Criteria 170, Assessing Human Health Risks of Chemicals: Derivation of Guidance Values for Health-Based Exposure Limits, IPCS International Programme on Chemical Safety; World Health Organization Geneva.

WHO, World Health Organization. (2005). Guidance Document for the Use of Data in Development of Chemical-Specific Adjustment Factors (CSAFs) for Interspecies Differences and Human Variability in Dose/Concentration-Response Assessment. Harmonization Project Document No. 2, World Health Organization, Geneva http://whqlibdoc.who.int/publications/2005/9241546786_eng.pdf.

WHO, World Health Organization. (2014). Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. IPCS Harmonization Project Document 11, WHO, Geneva, Switzerland. http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf.

Wiecek, W., Dorne, J.-L., Quignot, N., Bechaux, C., Amzal, B. (2019). A generic Bayesian hierarchical model for the meta-analysis of human population variability in kinetics and its applications in chemical risk assessment. Computational Toxicology 12, Article 100106.

Yang, X., Doerge, D.R., Teeguarden, J.G., Fisher, J.W. (2015). Development of a physiologically based pharmacokinetic model for assessment of human exposure to bisphenol A. Toxicology and Applied Pharmacology 289, 442-456.

Zeise, L., Bois, F.Y., Chiu, W.A., Hattis, D., Rusyn, I., Guyton, K.Z. (2013). Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environmental Health Perspectives 121, 23-31.

Zhang, B., Qi, S., Yue, Y., Shen, J., Li, C., Qian, W., Wu, J. (2018). Particle disposition in the realistic airway tree models of subjects with tracheal bronchus and COPD. BioMed Research International 2018, 7428609.

List of Tables

List of Figures

Annex 1 – Pubmed search strategies

History

Download history Clear history

Search	Add to builder	Query	Items found	Time
#21	Add	Search **(#13 or #20)** Filters: **published in the last 10 years**	240	04:18:50
#20	Add	Search **(#18 not #19)** Filters: **published in the last 10 years**	180	04:18:22
#18	Add	Search **(#16 and #17)** Filters: **published in the last 10 years**	237	04:17:50
#19	Add	Search **(#16 and #17)** Filters: **Review; published in the last 10 years**	57	04:17:44
#17	Add	Search **(interindividual or intraindividual or intersubject or intrasubject or between subject* or between individual*)** Filters: **published in the last 10 years**	27071	04:16:22
#16	Add	Search **(#15 and (dose or dosing or dosed))** Filters: **published in the last 10 years**	853	04:15:47
#15	Add	Search **(#14 and #11)** Filters: **published in the last 10 years**	1626	04:13:52
#14	Add	Search **(pharmacodynamic*[Title/Abstract] OR toxicodynamic*[Title/Abstract])** Filters: **published in the last 10 years**	19394	04:13:39
#13	Add	Search **(#10 and #11 and #12)** Filters: **published in the last 10 years**	60	04:05:07
#12	Add	Search **(auc[Title/Abstract] OR (area[Title/Abstract] AND under[Title/Abstract] AND curve) [Title/Abstract] OR (plasma[Title/Abstract] AND (concentration*[Title/Abstract] OR level*)) [Title/Abstract] OR cmax[Title/Abstract])** Filters: **published in the last 10 years**	5395	04:04:45
#11	Add	Search **((human or humans or volunteers or subjects or interindividual or intraindividual or intersubject or intrasubject) and (variability or variabilities or variation))** Filters: **published in the last 10 years**	215443	04:03:03
#10	Add	Search **((((genotype* or polymorph* or allele* or phenotype*) and (pharmacokinetic* or toxicokinetic* or metabolite* or metabolism))) AND "last 10 years"[PDat])** Filters: **published in the last 10 years**	229155	03:57:33

History

Download history Clear history

Search	Add to builder	Query	Items found	Time
#15	Add	Search (intraspecies or inter-individual variability or individual difference* individual variation or inter-individual variation) and (inhalative or inhalation or aerosol or dust) Filters: Toxicology	22	09:40:27
#105	Add	Search (#88 or #73 or #49 or #16 or #15) Filters: published in the last 10 years; Humans; Toxicology	97	09:40:27
#49	Add	Search ((inhalative or inhalation or aerosol or dust) and controlled exposure*) Filters: published in the last 10 years; Humans; Toxicology	35	09:40:27
#16	Add	Search (intraspecies or inter-individual variability) and (inhalative or inhalation or aerosol or dust)) Filters: Toxicology	48	09:40:27
#88	Add	Search (inhalative or inhalation or aerosol or dust) and controlled and (concentration or concentrations) and (variability or variation or inter-individual or individual difference*) Filters: published in the last 10 years; Humans; Toxicology	47	09:23:49
#87	Add	Search (inhalative or inhalation or aerosol or dust) and controlled and (concentration or concentrations) Filters: published in the last 10 years; Humans; Toxicology	643	08:49:44
#84	Add	Search (inhalative or inhalation or aerosol or dust) and controlled and concentration Filters: published in the last 10 years; Humans; Toxicology	437	08:48:51
#83	Add	Search (inhalative or inhalation or aerosol or dust) and controlled Filters: published in the last 10 years; Humans; Toxicology	2377	08:48:43
#73	Add	Search ((inhalative or inhalation or aerosol or dust) and controlled human and concentrations) and (variability or variation or interindividual difference* or individual difference*) Filters: published in the last 10 years; Humans; Toxicology	25	08:44:32
#69	Add	Search ((inhalative or inhalation or aerosol or dust) and controlled human and concentrations) Filters: published in the last 10 years; Humans; Toxicology	417	08:42:10
#67	Add	Search ((inhalative or inhalation or aerosol or dust) and controlled human) Filters: published in the last 10 years; Humans; Toxicology	2377	08:41:06
#30	Add	Search ((intraspecies or inter-individual variability or individual difference* or individual variation or inter-individual variation) and (inhalative or inhalation or aerosol or dust)) Filters: published in the last 10 years; Humans; Toxicology	211	08:29:08
#20	Add	Search ((intraspecies or inter-individual variability or individual difference* or individual variation or inter-individual variation) and (inhalative or inhalation or aerosol or dust)) Filters: Toxicology	694	07:59:38
#19	Add	Search ((intraspecies or inter-individual variability or individual difference* or individual variation or inter-individual variation) and (inhalative or inhalation or aerosol or dust))	1617	07:42:02

History Download history Clear history

Search	Add to builder	Query	Items found	Time
#5	Add	Search (((((genotype* OR polymorph* OR allele* OR phenotype*) AND (pharmacokinetic* OR toxicokinetic* OR metabolite* OR metabolism)) AND (auc OR (area and under and curve) OR (plasma AND (concentration* OR level*)) OR cmax) AND ((human OR humans OR volunteers OR subjects OR interindividual OR intraindividual OR intersubject OR intrasubject) AND (variability OR variabilities OR variation))) AND ("last 10 years"[PDat])) AND ((auc[Title/Abstract] OR (area[Title/Abstract] AND under [Title/Abstract] AND curve) AND Title/Abstract OR (plasma[Title/Abstract] AND (concentration* [Title/Abstract] OR level*)) AND Title/Abstract OR cmax[Title/Abstract]) AND ("last 10 years"[PDat]))) Filters: published in the last 10 years	63	09:19:29

History Download history Clear history

Search	Add to builder	Query	Items found	Time
#2	Add	Search ((((((interindividual OR intraindividual OR intersubject OR intrasubject OR between subject* OR between individual*) AND ("last 10 years"[PDat])) AND ((((((pharmacodynamic*[Title/Abstract] OR toxicodynamic*[Title/Abstract]) AND ("last 10 years"[PDat])) AND ((human OR humans OR volunteers OR subjects OR interindividual OR intraindividual OR intersubject OR intrasubject) AND (variability OR variabilities OR variation))) AND ("last 10 years"[PDat])) AND (dose OR dosing OR dosed)) AND ("last 10 years"[PDat]))) AND ("last 10 years"[PDat])) NOT ((((interindividual OR intraindividual OR intersubject OR intrasubject OR between subject* OR between individual*) AND ("last 10 years"[PDat])) AND ((((((pharmacodynamic*[Title/Abstract] OR toxicodynamic*[Title/Abstract]) AND ("last 10 years"[PDat])) AND ((human OR humans OR volunteers OR subjects OR interindividual OR intraindividual OR intersubject OR intrasubject) AND (variability OR variabilities OR variation))) AND ("last 10 years"[PDat])) AND (dose OR dosing OR dosed)) AND ("last 10 years"[PDat]))) AND (Review [ptyp] AND "last 10 years"[PDat]))) AND "last 10 years"[PDat]) Filters: published in the last 10 years	191	09:22:51

History Download history Clear history

Search	Add to builder	Query	Items found	Time
#2	Add	Search (((((genotype* OR polymorph* OR allele* OR phenotype*) AND (pharmacokinetic* OR toxicokinetic* OR metabolite* OR metabolism)) AND (auc OR (area and under and curve) OR (plasma and AND (concentration* OR level*)) OR cmax) AND ((human OR humans OR volunteers OR subjects OR interindividual OR intraindividual OR intersubject OR intrasubject) AND (variability OR variabilities OR variation))) AND ("last 10 years"[PDat])) AND ((auc[Title/Abstract] OR (area[Title/Abstract] AND under[Title/Abstract] AND curve) AND Title/Abstract OR (plasma[Title/Abstract] AND (concentration* [Title/Abstract] OR level*)) AND Title/Abstract OR cmax[Title/Abstract]) AND ("last 10 years"[PDat]))) Filters: published in the last 10 years	63	09:24:27

History Download history Clear history

Search	Add to builder	Query	Items found	Time
#1	Add	Search ((pharmacodynamic[Title/Abstract] OR pharmacodynamics[Title/Abstract] OR toxicodynamic [Title/Abstract] OR toxicodynamics[Title/Abstract])) AND (((((pharmacodynamic OR pharmacodynamics OR toxicodynamic OR toxicodynamics)) AND ((((human OR humans OR volunteers OR subjects OR interindividual OR intraindividual OR intersubject OR intrasubject) AND (variability OR variabilities OR variation)) AND ("last 10 years"[PDat]))) AND (pharmacological OR toxicological) AND (dose OR doses OR concentration OR concentrations) AND ("last 10 years"[PDat] AND Humans[Mesh])) AND ("last 10 years"[PDat] AND Humans[Mesh]))) AND (oral OR nasal OR oromucosal OR inhalation OR inhalative OR spray) AND ("last 10 years"[PDat] AND Humans[Mesh])) Filters: published in the last 10 years; Humans	135	09:25:36

History Download history Clear history

Search	Add to builder	Query	Items found	Time
#1	Add	Search ((((clinical OR volunteers OR subjects OR interindividual OR intraindividual OR intersubject OR intrasubject) AND (variability OR variabilities OR variation)) AND ("last 10 years"[PDat] AND Humans [Mesh]))) AND (mg/kg OR mg OR mg/m3) Filters: published in the last 10 years; Humans	388	09:26:58

Annex 2 – The concept of log GSD

For assessing and describing inter-individual differences in susceptibility the spread (or width) of distributions of individual values is important. For example, AUCs measured in a group of volunteers, who all have received the same dose. As such parameters cannot assume negative values, in most cases they are not normally distributed: they typically have a bound on the left side (zero) and are skewed to the right. As an approximation it can be assumed that they are lognormally distributed.

Some authors calculate ratios of percentiles divided by the median or average in order to characterise the spread of the distribution. However, as the distance between the e.g. 5^{th} percentile to the median is different to the distance between the median and the 95^{th} percentile, this approach must be seen critical.

For the same reason the width of lognormally distributed data cannot be described by their standard deviation either. However, when they are logarithmised a normal distribution is obtained, which can be characterised by mean and standard deviation. The standard deviation of the logarithmised values of a lognormal distribution here is called ”log GSD”, in analogy to WHO (2014). Further, the parameter GSD can be obtained from log GSD by exponentiation:
$GSD = 10^{\log GSD}$

According to WHO (2014), “GSD is a measure for the spread of a distribution, which is preferred over the standard deviation in case of lognormal or other right-sided distributions”. It can be used to compare the spread of distributions. But please note that GSD has no statistical definition in a sense that it represents the standard deviation of lognormally distributed values.

In our report we use log GSD to characterise the variability in datasets assumed to be lognormally distributed. For example, a distribution of log GSD values is obtained from the database on toxicokinetic data evaluated. From the distribution of log GSD values distributions of extrapolation factors can be developed.

For a given log GSD (i.e. for a given variability) the factor required to cover susceptibilities higher than the median can be calculated according to the following equation (WHO 2014):

Factor covering (1 – I) of the population = GSD^{z1-I}

where I is the incidence and z is the z-Score[15] corresponding to this incidence. For *I* = 5% the corresponding value for z_{1-I} is 1.6449, for I = 1% it is 2.3263.

[15] The z-score is a measure of how many standard deviations below or above the mean a value is: a z-score of 1 is 1 standard deviation above the mean (representing the 84^{th} percentile of the distribution), a z-score of 2 is 2 standard deviations above the mean (at the 97.5^{th} percentile of the distribution); the chosen z-scores of 1.6449 and 2.3263 represent the 95^{th} and 99^{th} percentile of the normal distribution

The following table shows, for orientation, the factors required to cover 95% of the population for various given values of log GSD, i.e. the factors covering the distance between the median (susceptibility) of the considered population and the 95th percentile.

Table A-1 Translation of log GSD into coefficient of variation (CV) and factors for covering 95% of the population (adapted from WHO (2014)).

Log GSD	**CV (%)**	**GSD (=10$^{log\ GSD}$)**	**Factor for 95% of population (=GSD$^{1.6449}$)***
0.1	23.3	1.259	1.46
0.2	48.6	1.585	2.13
0.3	78.2	1.995	3.11
0.4	116	2.512	4.55

*Calculated according to algorithm in WHO (2014) (see text)

Note that log GSD itself is subject to uncertainty and is described as a distribution. So, the median of the obtained log GSD distribution translates into a factor to cover the defined fraction of the population, e.g. 95%, with a 50% probability, whereas the 80th percentile of that log GSD distribution covers the 95% of the population with a probability of 80%.

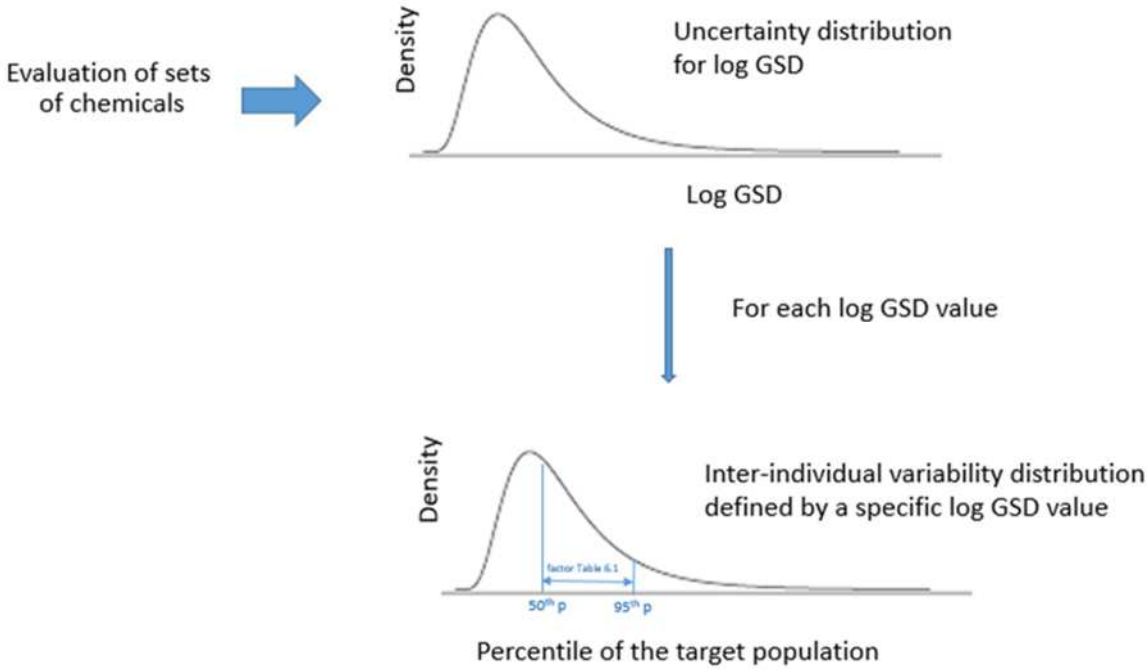


Figure A-1 Graphical representation of the relationship between the distribution of log GSD and the susceptibility distribution in the target population (adapted from WHO (2014))

For further illustration we use the result from our dataset on toxicokinetic differences (see **Table 2-2**) to calculate the factors required to cover 95% or 99% of the population with a probability of 50%, 75% or 95%.

Table A-2 Translation of various percentile values of the log GSD distribution for toxicokinetic data into factors for covering 95% or 99% of the population, respectively.

Percentiles of log GSD distribution	**Log GSD value for the selected percentile**	**GSD**	**Factor for 95% of population (=GSD$^{1.6449}$)***	**Factor for 99% of population (=GSD$^{2.3263}$)***
Median (50th percentile)	0.146	1.40	1.74	2.19
75th percentile	0.220	1.66	2.30	3.25
95th percentile	0.355	2.26	3.84	6.70

*Calculated according to algorithm in WHO (2014) (see above)

Annex 3 – Tabular summary of evaluated datasets

Toxicokinetic data

Reference	route	Substance class	healthy	n	Log GSD
Ahmed et al. 2015	oral	pharmaceutical	yes	9	0.351
Åkesson et al. 1988	inhalation	industrial	yes	5	0.060
Åkesson et al. 2000	inhalation	industrial	yes	6	NA[&]
Ali et al. 2012	oral	pharmaceutical	yes	8	0.233
Bastami et al. 2014	oral	pharmaceutical	yes	9	NA[&]
Breuer et al. 2013	inhalation	pharmaceutical	yes	8	0.180
Cho et al. 2009	oral	industrial	yes	8	0.128
Cho et al. 2011	oral	pharmaceutical	yes	46	0.171
Cho et al. 2018	oral	pharmaceutical	yes	6	0.142
Cook et al. 1991	inhalation	industrial	yes	12	0.106
Dimatteo et al. 2016	oral	pharmaceutical	no	92	0.264
Djebli et al. 2015	oral	pharmaceutical	yes	63	0.167
Dunbar et al. 2006	oral	pharmaceutical	yes	6	0.446
Eisenberg et al. 2014	inhalation	pharmaceutical	no	8	0.139
Elcombe et al. 2013	oral	pharmaceutical	no	6	0.091
Eldon et al. 2015	oral	pharmaceutical	yes	6	0.183
Ernstgård et al. 2003	inhalation	industrial	yes	8	0.274
Ernstgård et al. 2005	inhalation	industrial	yes	8	0.089
Falk-Filipsson et al. 1993	inhalation	industrial	yes	8	0.080
Falk et al. 1991	inhalation	industrial	yes	8	0.028
Falk Filipsson et al. 1990	inhalation	industrial	yes	8	0.114
Falk Filipsson et al. 1996	inhalation	industrial	yes	8	0.036
Fotoohi et al. 2016	oral	pharmaceutical	no	39	0.190
Groeseneken et al. 1986	inhalation	industrial	yes	5	0.179

Groeseneken et al. 1989	inhalation	industrial	yes	7	0.047
Haufroid et al. 2002	inhalation	industrial	yes	14	0.127
Hirawat et al. 2007	oral	pharmaceutical	yes	6	0.048
Hu et al. 2015	inhalation	pharmaceutical	yes	12	0.216
Jakubowski et al. 1987	inhalation	industrial	yes	5	0.277
Järnberg et al. 1998§	inhalation	industrial	yes	9	NA&
Järnberg et al. 1998§	inhalation	industrial	yes	8	NA&
Järvinen et al. 1999	inhalation	industrial	yes	4	0.104
Johannson et al. 2011	oral	pharmaceutical	yes	4	0.142
Kaddurah-Daouk et al. 2018	other	pharmaceutical	yes	43	NA&
Kharasch et al. 2015	oral	pharmaceutical	yes	44	0.362
Kostrzewski et al. 1997	inhalation	industrial	yes	5	0.075
Lacaillon et al. 1999	inhalation	pharmaceutical	yes	12	0.127
Lam et al. 1993	inhalation	industrial	yes	8	0.058
Larsby et al. 1986	inhalation	industrial	yes	10	0.219
Le Coutre et al. 2004	oral	pharmaceutical	no	6	0.261
Lee et al. 2009	oral	pharmaceutical	yes	6	0.062
Lin et al. 2001	inhalation	industrial	yes	133	0.175
Linn et al. 1995	inhalation	pharmaceutical	yes	7	0.102
Löf et al. 1993	inhalation	industrial	yes	9	0.092
Lund et al. 1997	inhalation	industrial	yes	7	0.227
Manitpisitkul et al. 2015	oral	pharmaceutical	yes	6	0.357
Meaklim et al. 2003§	oral	industrial	yes	12	0.405
Meaklim et al. 2003§	oral	industrial	yes	12	0.273
Mohanan et al. 2018	other	pharmaceutical	no	87	0.210
Mráz et al. 1989	inhalation	industrial	yes	10	0.231
Mráz et al. 1999§	inhalation	industrial	yes	8	0.060
Mráz et al. 1999§	inhalation	industrial	yes	8	0.077

Mráz et al. 1999[§]	inhalation	industrial	yes	8	0.086
Ogata et al. 2010	oral	pharmaceutical	yes	9	0.103
Okusanya et al. 2014	inhalation	pharmaceutical	no	105	NA[&]
Potocka et al. 2010	oral	pharmaceutical	yes	6	0.220
Rocha et al. 2012	oral	pharmaceutical	yes	6	0.175
Saiz-Rodríguez et al. 2018	other	pharmaceutical	yes	35	0.200
Sandquist et al. 2013	oral	pharmaceutical	no	12	0.268
Schaffler et al. 2013	oral	pharmaceutical	yes	24	0.112
Schnell et al. 2014	oral	pharmaceutical	yes	8	0.097
Small et al. 2010	oral	pharmaceutical	yes	22	0.150
Stahlbom et al. 1991	inhalation	industrial	yes	4	0.154
Stangier et al. 2007	oral	pharmaceutical	yes	8	0.142
Stass et al. 2013	inhalation	pharmaceutical	no	6	0.190
Stott et al. 2013	other	pharmaceutical	yes	7	0.317
Stott et al. 2018	oral	pharmaceutical	no	43	0.052
van Staveren et al. 2011	oral	pharmaceutical	yes	11	0.120
van Staveren et al. 2012	oral	pharmaceutical	yes	11	0.136
Vanky et al. 2017	other	pharmaceutical	yes	20	0.169
Venail et al. 2018	oral	pharmaceutical	yes	9	0.109
Wenker et al. 2001	inhalation	industrial	yes	20	0.091
Zhang et al. 2018	oral	pharmaceutical	yes	38	0.287
Zimmerman et al. 1999	oral	pharmaceutical	yes	23	0.149

[§] = different substance or tested condition in the same study

[&] = the parameters used to describe the distribution obtained in this study could not be used to derive a log GSD value

Toxicodynamic data

Reference	route	Substance class	healthy	n	Dose ratio
Adefurin et at. 2017	parenteral	pharmaceutical	yes	41	5
Andersen et al. 1974	inhalation	industrial	yes	15	6
Beck, 1959 as cited in WHO 1982	inhalation	industrial	no	10	7
Bryson et al. 2018	other	pharmaceutical	no	500	8
Falvella et al. 2016	other	pharmaceutical	no	53	16
Heerdt et al. 2016	parenteral	pharmaceutical	yes	34	3
Hine et al. 1960§	inhalation	industrial	yes	35	36
Hine et al. 1960§	inhalation	industrial	yes	35	201
Hirawat et al. 2007	oral	pharmaceutical	yes	6	3
Industrial Bio-Test Lab., as cited in Greim 1993	inhalation	industrial	yes	10	5
Kulle et al. 1993	inhalation	industrial	yes	19	7
Lund et al. 1997	inhalation	industrial	yes	20	4
Lund et al. 1999	inhalation	industrial	yes	19	8
Macdonald et al. 1993	oral	pharmaceutical	yes	18	4
Monse et al. 2018	inhalation	industrial	yes	16	5
Muszkat et al. 2011§	parenteral	pharmaceutical	yes	62	14
Muszkat et al. 2011§	parenteral	pharmaceutical	yes	62	11
Roger et al. 1985	inhalation	industrial	no	28	5
Sheppard et al. 1980	inhalation	industrial	yes	21	6
Shusterman et al. 1997 and 2003	inhalation	industrial	yes	20/60	4
Sigurjónsdóttir et al. 2001	other	other	yes	Between 10 to 30	8
Chriguer et al. 2005	oral	pharmaceutical	yes	40	5
Stangier et al. 2007	oral	pharmaceutical	yes	40	3
Zhang et al. 2016	parenteral	pharmaceutical	no	27	8

§ = different substances or tested condition in the same study

References database

Toxicokinetik data

Ahmed, A.I.; van den Elsen, G.A.; Colbers, A.; Kramers, C.; Burger, D.M.; van der Marck, M.A.; Olde Rikkert, M.G. (2015): Safety, pharmacodynamics, and pharmacokinetics of multiple oral doses of delta-9-tetrahydrocannabinol in older persons with dementia. Psychopharmacology (Berl), 232, 2587-2595

Åkesson, B.; Jonsson, B.A. (2000): Biological monitoring of N-methyl-2-pyrrolidone using 5-hydroxy-N-methyl-2-pyrrolidone in plasma and urine as the biomarker. Scandinavian Journal of Work, Environment and Health, 26, 213-218

Åkesson, B.; Skerfving, S.; Mattiasson, L. (1988): Experimental study on the metabolism of triethylamine in man. British Journal of Industrial Medicine, 45, 262-268

Ali, Z.; Laurijssens, B.; Ostenfeld, T.; McHugh, S.; Stylianou, A.; Scott-Stevens, P.; Hosking, L.; Dewit, O.; Richardson, J.C.; Chen, C. (2013): Pharmacokinetic and pharmacodynamic profiling of a P2X7 receptor allosteric modulator GSK1482160 in healthy human subjects. British Journal of Clinical Pharmacology, 75, 197-207

Bastami, S.; Haage, P.; Kronstrand, R.; Kugelberg, F.C.; Zackrisson, A.L.; Uppugunduri, S. (2014): Pharmacogenetic aspects of tramadol pharmacokinetics and pharmacodynamics after a single oral dose. Forensic Science International, 238, 125-132

Breuer, C.; Wachall, B.; Gerbeth, K.; Abdel-Tawab, M.; Fuhr, U. (2013): Pharmacokinetics and pharmacodynamics of moist inhalation epinephrine using a mobile inhaler. European Journal of Clinical Pharmacology, 69, 1303-1310

Cho, H.J.; Lee, S.Y.; Kim, Y.G.; Oh, S.Y.; Kim, J.W.; Huh, W.; Ko, J.W.; Kim, H.G. (2011): Effect of genetic polymorphisms on the pharmacokinetics and efficacy of glimepiride in a Korean population. Clinica Chimica Acta, 412, 1831-1834

Cho, H.Y.; Blum, R.A.; Sunderland, T.; Cooper, G.J.; Jusko, W.J. (2009): Pharmacokinetic and pharmacodynamic modeling of a copper-selective chelator (TETA) in healthy adults. Journal of Clinical Pharmacology, 49, 916-928

Cho, Y.S.; Lim, H.S.; Cho, Y.L.; Nam, H.S.; Bae, K.S. (2018): Multiple-dose safety, tolerability, pharmacokinetics, and pharmacodynamics of oral LCB01-0371 in healthy male volunteers. Clinical Therapeutics, 40, 2050-2064

Cook, M.R.; Bergman, F.J.; Cohen, H.D.; Gerkovich, M.M.; Graham, C.; Harris, R.K.; Siemann, L.G. (1991): Effects of methanol vapor on human neurobehavioral measures. Research Report / Health Effects Institute (Cambridge), 1-45

Dimatteo, C.; D'Andrea, G.; Vecchione, G.; Paoletti, O.; Cappucci, F.; Tiscia, G.L.; Buono, M.; Grandone, E.; Testa, S.; Margaglione, M. (2016): Pharmacogenetics of dabigatran etexilate interindividual variability. Thrombosis Research, 144, 1-5

Djebli, N.; Fabre, D.; Boulenc, X.; Fabre, G.; Sultan, E.; Hurbin, F. (2015) Physiologically based pharmacokinetic modeling for sequential metabolism: effect of CYP2C19 genetic polymorphism on clopidogrel and clopidogrel active metabolite pharmacokinetics

Drug Metabolism and Disposition, 43, 510 Dunbar, G.; Demazières, A.; Monreal, A.; Cisterni, C.; Metzger, D.; Kuchibhatla, R.; Luthringer, R. (2006): Pharmacokinetics and safety profile of ispronicline (TC-1734), a new brain nicotinic receptor partial agonist, in young healthy male volunteers. Journal of Clinical Pharmacology, 46, 715-726

Eisenberg, E.; Ogintz, M.; Almog, S. (2014): The pharmacokinetics, efficacy, safety, and ease of use of a novel portable metered-dose cannabis inhaler in patients with chronic neuropathic pain: a phase 1a study. Journal of Pain & Palliative Care Pharmacotherapy, 28, 216-225

Elcombe, C.R.; Wolf, C.R.; Westwood, D. (2013): United States Patent Application Publication. Compositions comprising perfluorooctanoic acid. Pub. No. US 2013/0029928 https://patentimages.storage.googleapis.com/24/ee/73/f58267c7d70dde/WO2011101643A1.pdf

Eldon, M.A.; Kugler, A.R.; Medve, R.A.; Bui, K.; Butler, K.; Sostek, M. (2015): Safety, tolerability, pharmacokinetics, and pharmacodynamic effects of naloxegol at peripheral and central nervous system receptors in healthy male subjects: A single ascending-dose study. Clinical Pharmacology in Drug Development, 4, 434-441

Ernstgård, L.; Shibata, E.; Johanson, G. (2005): Uptake and disposition of inhaled methanol vapor in humans. Toxicological Sciences, 88, 30-38

Ernstgård, L.; Sjögren, B.; Warholm, M.; Johanson, G. (2003): Sex differences in the toxicokinetics of inhaled solvent vaporsin humans 1. m-Xylene. Toxicology and Applied Pharmacology, 193, 147-157

Falk, A.; Löf, A.; Hagberg, M.; Hjelm, E.W.; Wang, Z. (1991): Human exposure to 3-carene by inhalation: toxicokinetics, effects on pulmonary function and occurrence of irritative and CNS symptoms. Toxicology and Applied Pharmacology, 110, 198-205

Falk, A.A.; Hagberg, M.T.; Loef, A.E.; Wigaeus-Hjelm, E.M.; Zhiping, W. (1990): Uptake, distribution and elimination of alpha-pinene in man after exposure by inhalation. Scandinavian Journal of Work, Environment and Health, 16, 372-378

Falk-Filipsson, A. (1996): Short term inhalation exposure to turpentine: toxicokinetics and acute effects in men. Occupational and Environmental Medicine, 53, 100-106

Falk-Filipsson, A.; Lof, A.; Hagberg, M.; Hjelm, E.W.; Wang, Z. (1993): d-limonene exposure to humans by inhalation: uptake, distribution, elimination, and effects on the pulmonary function. Journal of Toxicology and Environmental Health, 38, 77-88

Fotoohi, A.K.; Karim, H.; Lafolie, P.; Pohanka, A.; Östervall, J.; Hatschek, T.; Vitols, S. (2016): Pronounced interindividual but not intraindividual variation in tamoxifen and metabolite levels in plasma during adjuvant treatment of women with early breast cancer. Therapeutic Drug Monitoring, 38, 239-245

Groeseneken, D.; van Vlem, E.; Veulemans, H.; Masschelein, R. (1986): Gas chromatographic determination of methoxyacetic and ethoxyacetic acid in urine. British Journal of Industrial Medicine, 43, 62-65

Groeseneken, D.; Veulemans, H.; Masschelein, R.; van Vlem, E. (1989): Experimental human exposure to ethylene glycol monomethyl ether. International Archives of Occupational and Environmental Health, 61, 243-247

Haufroid, V.; Jakubowski, M.; Janasik, B.; Ligocka, D.; Buchet, J.P.; Bergamaschi, E.; Manini, P.; Mutti, A.; Ghittori, S.; Arand, M.; Hangen, N.; Oesch, F.; Hirvonen, A.; Lison, D. (2002): Interest of genotyping and phenotyping of drug-metabolizing enzymes for the interpretation of biological monitoring of exposure to styrene. Pharmacogenetics, 12, 691-702

Hirawat, S.; Welch, E.M.; Elfring, G.L.; Northcutt, V.J.; Paushkin, S.; Hwang, S.; Leonard, E.M.; Almstead, N.G.; Ju, W.; Peltz, S.W.; Miller, L.L. (2007): Safety, tolerability, and pharmacokinetics of PTC124, a nonaminoglycoside nonsense mutation suppressor, following single- and multiple-dose administration to healthy male and female adult volunteers. Journal of Clinical Pharmacology, 47, 430-444

Hu, C.; Jia, J.; Dong, K.; Luo, L.; Wu, K.; Mehta, R.; Peng, J.; Ren, Y.; Gross, A.; Yu, H. (2015): Pharmacokinetics and tolerability of inhaled umeclidinium and vilanterol alone and in combination in healthy Chinese subjects: a randomized, open-label, crossover trial. PLoS One, 10, e0121264

Jakubowski, M.; Linhart, I.; Pielas, G.; Kopecky, J. (1987): 2-Cyanoethylmercapturic acid (CEMA) in the urine as a possible indicator of exposure to acrylonitrile. British Journal of Industrial Medicine, 44, 834-840

Järnberg, J.; Johanson, G.; Löf, A.; Stahlbom, B. (1998): Toxicokinetics of 1,2,4-trimethylbenzene in humans exposed to vapours of white spirit: comparison with exposure to 1,2,4-trimethylbenzene alone. Archives of Toxicology, 72, 483-491

Järvinen, P.; Engström, K.; Riihimäki, V.; Ruusuvaara, P.; Setälä, K. (1999): Effects of experimental exposure to triethylamine on vision and the eye. Occupational and Environmental Medicine, 65, 1-5

Johansson, S.; Cullberg, M.; Eriksson, U.G.; Elg, M.; Dunér, K.; Jensen, E.; Wollbratt, M.; Wåhlander, K. (2011): Single-dose pharmacokinetics, pharmacodynamics and safety of AZD0837, a novel oral direct thrombin inhibitor, in young healthy male subjects. International Journal of Clinical Pharmacology and Therapeutics, 49, 258-267

Kaddurah-Daouk, R.; Hankemeier, T.; Scholl, E.H.; Baillie, R.; Harms, A.; Stage, C.; Dalhoff, K.P.; Jűrgens, G.; Taboureau, O.; Nzabonimpa, G.S.; Motsinger-Reif, A.A.; Thomsen, R.; Linnet, K.; Rasmussen, H.B.; INDICES Consortium; Pharmacometabolomics Research Network (2018): Pharmacometabolomics informs about pharmacokinetic profile of methylphenidate. CPT: pharmacometrics & systems pharmacology, 7, 525-533

Kharasch, E.D.; Regina, K.J.; Blood, J.; Friedel, C. (2015): Methadone pharmacogenetics: CYP2B6 polymorphisms determine plasma concentrations, clearance, and metabolism. Anesthesiology, 123, 1142-1153

Kostrzewski, P.; Wiaderna-Brycht, A.; Czerski, B. (1997): Biological monitoring of experimental human exposure to trimethylbenzene. The Science of the Total Environment, 199, 73-81

Lam, C.W.; Weir, F.W.; Williams-Cavender, K.; Tan, M.N.; Galen, T.J.; Pierson, D.L. (1993): Toxicokinetics of inhaled bromotrifluoromethane (Halon 1301) in human subjects. Fundamental and Applied Toxicology, 20, 231-239

Larsby, B.; Tham, R.; Eriksson, B.; Hydén, D.; Ödkvist, L.; Liedgren, C.; Bunnfors, I. (1986): Effects of trichloroethylene on the human vestibulo-oculomotor system. Acta Oto-Laryngologica, 101, 193-199

le Coutre, P.; Kreuzer, K.A.; Pursche, S.; Bonin, M.; Leopold, T.; Baskaynak, G.; Dörken, B.; Ehninger, G.; Ottmann, O.; Jenke, A.; Bornhäuser, M.; Schleyer, E. (2004): Pharmacokinetics and cellular uptake of imatinib and its main metabolite CGP74588. Cancer Chemotherapy and Pharmacology, 53, 313-323

Lecaillon, J.B.; Kaiser, G.; Palmisano, M.; Morgan, J.; Della Cioppa, G. (1999): Pharmacokinetics and tolerability of formoterol in healthy volunteers after a single high dose of Foradil dry powder inhalation via aerolizerTM. European Journal of Clinical Pharmacology, 55, 131-138

Lee, L.; Niu, H.; Rueger, R.; Igawa, Y.; Deutsch, J.; Ishii, N.; Mu, S.; Sakamoto, Y.; Busse-Reid, R.; Gimmi, C.; Goelzer, P.; De Schepper, S.; Yoshimura, Y.; Barrett, J.; Ishikawa, Y.; Weissgerber, G.; Peck, R. (2009): The safety, tolerability, pharmacokinetics, and pharmacodynamics of single oral doses of CH4987655 in healthy volunteers: target suppression using a biomarker. Clin Cancer Res, 15, 7368-7374

Lin, Y.-S.; Smith, T.J.; Kelsey, K.T.; Wypij, D. (2001): Human physiologic factors in respiratory uptake of 1,3-butadiene. Environmental Health Perspectives, 109, 921-926

Linn, W.S.; Gong, H.; Anderson, K.R.; Clark, K.W.; Shamoo, D.A. (1995): Exposures of health-care workers to ribavirin aerosol: a pharmacokinetic study. Archives of Environmental Health, 50, 445-451

Löf, A.; Wigaeus Hjelm, E.; Colmsjö, A.; Lundmark, B.O.; Norström, A.; Sato, A. (1993): Toxicokinetics of toluene and urinary excretion of hippuric acid after human exposure to 2H8-toluene. British Journal of Industrial Medicine, 50, 55-59

Lund, K.; Ekstrand, J.; Boe, J.; Sostrand, P.; Kongerud, J. (1997): Exposure to hydrogen fluoride: an experimental study in humans of concentrations of fluoride in plasma, symptoms, and lung function. Occupational and Environmental Medicine, 54, 32-37

Manitpisitkul, P.; Mayorga, A.; Shalayda, K.; De Meulder, M.; Romano, G.; Jun, C.; Moyer, J.A. (2015): Safety, tolerability and pharmacokinetic and pharmacodynamic learnings from a double-blind, randomized, placebo-controlled, sequential group first-in-human study of the TRPV1 antagonist, JNJ-38893777, in healthy men. Clinical Drug Investigation, 35, 353-363

Meaklim, J.; Yang, J.; Drummer, O.H.; Killalea, S.; Staikos, V.; Horomidis, S.; Rutherford, D.; Ioannides-Demos, L.L.; Lim, S.; McLean, A.J.; McNeil, J.J. (2003): Fenitrothion: toxicokinetics and toxicologic evaluation in human volunteers. Environmental Health Perspectives, 111, 305-308

Mohanan, E.; Panetta, J.C.; Lakshmi, K.M.; Edison, E.S.; Korula, A.; Na, F.; Abraham, A.; Viswabandya, A.; George, B.; Mathews, V.; Srivastava, A.; Balasubramanian, P. (2018): Pharmacokinetics and pharmacodynamics of treosulfan in patients with thalassemia major undergoing allogeneic hematopoietic stem cell transplantation. Clinical Pharmacology and Therapeutics, 104, 575-583

Mráz, J.; Cross, H.; Gescher, A.; Threadgill, M.D.; Flek, J. (1989): Differences between rodents and humans in the metabolic toxification of N,N-dimethylformamide. Toxicology and Applied Pharmacology, 98, 507-516

Mráz, J.; Gálová, E.; Nohová, H.; Vítková, D.; Tichý, M. (1999): Effect of ethanol on the urinary excretion of cyclohexanol and cyclohexanediols, biomarkers of the exposure to cyclohexanone, cyclohexane and cyclohexanol in humans. Scandinavian Journal of Work, Environment and Health, 25, 233-237

Ogata, K.; Mendell-Harary, J.; Tachibana, M.; Masumoto, H.; Oguma, T.; Kojima, M.; Kunitada, S. (2010): Clinical safety, tolerability, pharmacokinetics, and pharmacodynamics of the novel factor Xa inhibitor edoxaban in healthy volunteers. Journal of Clinical Pharmacology, 50, 743-753

Okusanya, O.O.; Bhavnani, S.M.; Hammel, J.P.; Forrest, A.; Bulik, C.C.; Ambrose, P.G.; Gupta, R. (2014): Evaluation of the pharmacokinetics and pharmacodynamics of liposomal amikacin for inhalation in cystic fibrosis patients with chronic pseudomonal infections using data from two phase 2 clinical studies. Antimicrobial Agents and Chemotherapy, 58, 5005-5015

Potocka, E.; Cassidy, J.P.; Haworth, P.; Heuman, D.; van Marle, S.; Baughman, R.A., Jr. (2010): Pharmacokinetic characterization of the novel pulmonary delivery excipient fumaryl diketopiperazine. Journal of Diabetes Science and Technology, 4, 1164-1173

Rocha, J.F.; Vaz-Da-Silva, M.; Nunes, T.; Igreja, B.; Loureiro, A.I.; Bonifácio, M.J.; Wright, L.C.; Falcão, A.; Almeida, L.; Soares-Da-Silva, P. (2012): Single-dose tolerability, pharmacokinetics, and pharmacodynamics of etamicastat (BIA 5-453), a new dopamine b-hydroxylase inhibitor, in healthy subjects. Journal of Clinical Pharmacology, 52, 156-170

Saiz-Rodríguez, M.; Ochoa, D.; Herrador, C.; Belmonte, C.; Román, M.; Alday, E.; Koller, D.; Zubiaur, P.; Mejía, G.; Hernández-Martínez, M.; Abad-Santos, F. (2019): Polymorphisms associated with fentanyl pharmacokinetics, pharmacodynamics and adverse effects. Basic & Clinical Pharmacology & Toxicology, 124, 321-329

Sandqvist, A.M.; Henrohn, D.; Schneede, J.; Hedeland, M.; Egeröd, H.C.; Bondesson, U.G.; Wikström, B.G. (2013): High inter-individual variability of vardenafil pharmacokinetics in patients with pulmonary hypertension. European Journal of Clinical Pharmacology, 69, 197-207

Schaffler, K.; Reeh, P.; Duan, W.R.; Best, A.E.; Othman, A.A.; Faltynek, C.R.; Locke, C.; Nothaft, W. (2013): An oral TRPV1 antagonist attenuates laser radiant-heat-evoked potentials and pain ratings from UV(B)-inflamed and normal skin. British Journal of Clinical Pharmacology, 75, 404-414

Schnell, D.; Buschke, S.; Fuchs, H.; Gansser, D.; Goeldner, R.G.; Uttenreuther-Fischer, M.; Stopfer, P.; Wind, S.; Petersen-Sylla, M.; Halabi, A.; Koenen, R. (2014): Pharmacokinetics of afatinib in subjects with mild or moderate hepatic impairment. Cancer Chemotherapy and Pharmacology, 74, 267-275

Small, D.S.; Kothare, P.; Yuen, E.; Lachno, D.R.; Li, Y.G.; Winters, K.J.; Farid, N.A.; Ni, L.; Jakubowski, J.A.; Salazar, D.E.; Thieu, V.T.; Payne, C.D. (2010): The pharmacokinetics and pharmacodynamics of prasugrel in healthy Chinese, Japanese, and Korean subjects compared with healthy Caucasian subjects. European Journal of Clinical Pharmacology, 66, 127-135

Stahlbom, B.; Lundh, T.; Åkesson, B. (1991): Experimental study on the metabolism of dimethylethylamine in man. International Archives of Occupational and Environmental Health, 63, 305-310

Stangier, J.; Rathgen, K.; Stahle, H.; Gansser, D.; Roth, W. (2007): The pharmacokinetics, pharmacodynamics and tolerability of dabigatran etexilate, a new oral direct thrombin inhibitor, in healthy male subjects. British Journal of Clinical Pharmacology, 64, 292-303

Stass, H.; Weimann, B.; Nagelschmitz, J.; Rolinck-Werninghaus, C.; Staab, D. (2013): Tolerability and pharmacokinetic properties of ciprofloxacin dry powder for inhalation in patients with cystic fibrosis: a phase I, randomized, dose-escalation study. Clinical Therapeutics, 35, 1571-1581

Stott, C.G.; White, L.; Wright, S.; Wilbraham, D.; Guy, G.W. (2013): A phase I study to assess the single and multiple dose pharmacokinetics of THC/CBD oromucosal spray. European Journal of Clinical Pharmacology, 69, 1135-1147

Stott, K.E.; Beardsley, J.; Kolamunnage-Dona, R.; Castelazo, A.S.; Kibengo, F.M.; Mai, N.T.H.; Tùng, N.L.N.; Cuc, N.T.K.; Day, J.; Hope, W. (2018): Population pharmacokinetics and cerebrospinal fluid penetration of fluconazole in adults with cryptococcal meningitis. Antimicrobial Agents and Chemotherapy, 62, e00885-00818

van Staveren, M.C.; Theeuwes-Oonk, B.; Guchelaar, H.J.; van Kuilenburg, A.B.; Maring, J.G. (2011): Pharmacokinetics of orally administered uracil in healthy volunteers and in DPD-deficient patients, a possible tool for screening of DPD deficiency. Cancer Chemotherapy and Pharmacology, 68, 1611-1617

Vanky, E.; Hellmundt, L.; Bondesson, U.; Eksborg, S.; Lundeberg, S. (2017): Pharmacokinetics after a single dose of naloxone administered as a nasal spray in healthy volunteers. Acta Anaesthesiologica Scandinavica, 61, 636-640

Venail, F., et al. (2018). "Safety, tolerability, pharmacokinetics and pharmacokinetic-pharmacodynamic modelling of the novel H(4) receptor inhibitor SENS-111 using a modified caloric test in healthy subjects." British Journal of Clinical Pharmacology 84(12): 2836-2848.

Wenker, M.A.; Kezic, S.; Monster, A.C.; de Wolff, F.A. (2001): Metabolic capacity and interindividual variation in toxicokinetics of styrene in volunteers. Human and Experimental Toxicology, 20, 221-228

Zhang, H.; Li, Q.; Zhu, X.; Wu, M.; Li, C.; Li, X.; Liu, C.; Shen, Z.L.; Ding, Y.; Hua, S. (2018) Association of variability and pharmacogenomics with bioequivalence of gefitinib in healthy male subjects. Front Pharmacol, 9, 849-849

Zimmerman, J.J.; Ferron, G.M.; Lim, H.-K.; Parker, V.H. (1999) The effect of a high-fat meal on the oral bioavailability of the immunosuppressant Sirolimus (Rapamycin) The Journal of Clinical Pharmacology, 39, 1155-1161

Toxicodynamic data

Adefurin, A.; Ghimire, L.V.; Kohli, U.; Muszkat, M.; Sofowora, G.G.; Li, C.; Levinson, R.T.; Paranjape, S.Y.; Stein, C.M.; Kurnik, D. (2017): Genetic variation in the alpha1B-adrenergic receptor and vascular response. Pharmacogenomics Journal, 17, 366-371

Andersen, I.; Lundqvist, G.R.; Jensen, P.L.; Proctor, D.F. (1974) Human response to controlled levels of sulfur dioxide. Archives of Environmental Health, 28, 31-39

Bryson, E.O.; Kellner, C.H.; Li, E.H.; Aloysi, A.S.; Majeske, M. (2018): Extreme variability in succinylcholine dose for muscle relaxation in electroconvulsive therapy. Australasian Psychiatry, 26, 391-393

Falvella, F.S.; Marelli, S.; Cheli, S.; Montanelli, S.; Viecca, F.; Salvi, L.; Ferrara, A.; Clementi, E.; Trifiro, G.; Pini, A. (2016): Pharmacogenetic approach to losartan in Marfan patients: a starting point to improve dosing regimen? Drug Metabolism and Personalized Therapy, 31, 157-163

Greim, H. (1993). Gesundheitsschädliche Arbeitsstoffe, Toxikologisch-arbeitsmedizinische Begründungen von MAK-Werten, Loseblattsammlung, 19. Lfg. DFG Deutsche Forschungsgemeinschaft, VCH Verlag Weinheim and WHO, World Health Organization (1986). Environmental Health Criteria 54, Ammonia. IPCS International Programme on Chemical Safety; World Health Organization Geneva

Heerdt, P.M.; Sunaga, H.; Owen, J.S.; Murrell, M.T.; Malhotra, J.K.; Godfrey, D.; Steinkamp, M.; Savard, P.; Savarese, J.J.; Lien, C.A. (2016): Dose-response and cardiopulmonary side effects of the novel neuromuscular-blocking drug CW002 in man. Anesthesiology, 125, 1136-1143

Hine, C.H.; Kodama, J.K.; Guzman, R.J.; Loquvam, G.S. (1960). The toxicity of allylamines. Archives of Environmental Health, 1, 343-352

Hirawat, S.; Welch, E.M.; Elfring, G.L.; Northcutt, V.J.; Paushkin, S.; Hwang, S.; Leonard, E.M.; Almstead, N.G.; Ju, W.; Peltz, S.W.; Miller, L.L. (2007): Safety, tolerability, and pharmacokinetics of PTC124, a nonaminoglycoside nonsense mutation suppressor, following single- and multiple-dose administration to healthy male and female adult volunteers. Journal of Clinical Pharmacology, 47, 430-444

Kulle, T.J. (1993): Acute odor and irritation response in healthy nonsmokers with formaldehyde exposure. Inhalation Toxicology, 5, 323-332

Lund, K.; Ekstrand, J.; Boe, J.; Sostrand, P.; Kongerud, J. (1997): Exposure to hydrogen fluoride: an experimental study in humans of concentrations of fluoride in plasma, symptoms, and lung function. Occupational and Environmental Medicine, 54, 32-37

Lund, K.; Refsnes, M.; Sandstrom, T.; Sostrand, P.; Schwarze, P.; Boe, J.; Kongerud, J. (1999): Increased CD3 positive cells in bronchoalveolar lavage fluid after hydrogen fluoride inhalation. Scandinavian Journal of Work, Environment and Health, 25, 326-334

Macdonald, N.J.; Sioufi, A.; Howie, C.A.; Wade, J.R.; Elliott, H.L. (1993): The effects of age on the pharmacokinetics and pharmacodynamics of single oral doses of benazepril and enalapril. British Journal of Clinical Pharmacology, 36, 205-209

Monse, C.; Hagemeyer, O.; Raulf, M.; Jettkant, B.; van Kampen, V.; Kendzia, B.; Gering, V.; Kappert, G.; Weiss, T.; Ulrich, N.; Marek, E.M.; Bunger, J.; Bruning, T.; Merget, R. (2018): Concentration-dependent systemic response after inhalation of nano-sized zinc oxide particles in human volunteers. Particle and Fibre Toxicology, 15, 8

Muszkat, M.; Kurnik, D.; Sofowora, G.G.; Wood, A.J.; Stein, C.M. (2011): Independent regulation of a1 and a2 adrenergic receptor-mediated vasoconstriction in vivo. Journal of Hypertension, 29, 251-256

Roger, L.J.; Kehrl, H.R.; Hazucha, M.; Horstman, D.H. (1985): Bronchoconstriction in asthmatics exposed to sulfur dioxide during repeated exercise. Journal of Applied Physiology, 59, 784-791

Sheppard, D.; Wong, W.S.; Uehara, C.F.; Nadel, J.A.; Boushey, H.A. (1980) Lower threshold and greater bronchomotor responsive of asthmatic subjects to sulfur dioxideness American Review of Respiratory Diseases, 122, 873-878

Shusterman, D.; Balmes, J. (1997). Measurement of nasal irritant sensitivity to pulsed carbon dioxide: a pilot study. Archives of Environmental Health, 52, 334-340 AND Shusterman, D.; Murphy, M.A.; Balmes, J. (2003). Differences in nasal irritant sensitivity by age, gender, and allergic rhinitis status. International Archives of Occupational and Environmental Health, 76, 577-583

Sigurjónsdóttir, H.A.; Franzson, L.; Manhem, K.; Ragnarsson, J.; Sigurdsson, G.; Wallerstedt, S. (2001): Liquorice-induced rise in blood pressure: a linear dose-response relationship. Journal of Human Hypertension, 15, 549-552

Chriguer, R.S.; Elias, L.L.; da Silva, I.M., Jr.; Vieira, J.G.; Moreira, A.C.; de Castro, M. (2005): Glucocorticoid sensitivity in young healthy individuals: in vitro and in vivo studies. Journal of Clinical Endocrinology & Metabolism, 90, 5978-5984

Stangier, J.; Rathgen, K.; Stahle, H.; Gansser, D.; Roth, W. (2007): The pharmacokinetics, pharmacodynamics and tolerability of dabigatran etexilate, a new oral direct thrombin inhibitor, in healthy male subjects. British Journal of Clinical Pharmacology, 64, 292-303

WHO, World Health Organization (1982). Environmental Health Criteria 21, Chlorine and Hydrogen Chloride. IPCS International Programme on Chemical Safety; World Health Organization Geneva

Zhang, X.; Imel, E.A.; Ruppe, M.D.; Weber, T.J.; Klausner, M.A.; Ito, T.; Vergeire, M.; Humphrey, J.; Glorieux, F.H.; Portale, A.A.; Insogna, K.; Carpenter, T.O.; Peacock, M. (2016): Pharmacokinetics and pharmacodynamics of a human monoclonal anti-FGF23 antibody (KRN23) in the first multiple ascending-dose trial treating adults with X-linked hypophosphatemia. Journal of Clinical Pharmacology, 56, 176-185

REPORT 9:
Human Equivalent Concentration and Kinetic Modelling of Aerosols in the Lower Respiratory Tract

RESEARCH PROJECT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Content

Extended Summary

The “Human Equivalent Concentration” (HEC) approach is a procedure to extrapolate an exposure concentration from an experimental animal study to an equivalent human concentration for a chronic workplace inhalation exposure scenario. Within this report we discuss HEC calculations for solid particles in the lower respiratory tract. Recent developments of the HEC approach (scientific update; new calculation procedures; improvements; uncertainties) are described and compared to earlier versions. This analysis can be used by regulatory bodies to establish guidance on how to apply the HEC approach in regulatory procedures aiming at deriving occupational exposure limits (OEL) for particles affecting the lower respiratory tract.

A four step procedure

Exposure concentrations of particles in experimental animal studies are not regarded as equivalent to workplace concentrations due to several reasons: i) the intake into the lower respiratory tract depends on breathing patterns (nose or mouth breathing); also, breathing frequency and breathing volume differ significantly between rodents and humans, ii) the morphology of the human respiratory tract is different from that of rodents, which consequently leads to differences in deposition of particles in the lower respiratory tract including the lung; for example, a much higher fraction of respirable particles with a Mass Median Aerodynamic Diameter (MMAD) of more than 2 µm is deposited in the deep human lung compared to the rat lung; iii) once the particles reached the lung, the respective contact sites in the two species (usually rats and humans are compared) are highly different with regard to volume or surface area at the contact sites; defence or adverse responses or other biological reactions in the local lung environment will be initiated and in consequence interspecies differences of, e.g., the alveolar surface areas or the macrophage capacity may subsequently lead to different responses; iv) finally, translocation within the lung and elimination of particles from the lung have been observed to be highly different between many animal species and humans.

HEC aims at correcting the external concentration for differences between animals and humans with regard to the retained dose of particles in the lung. Consequently, a multi-step procedure is taken: four interspecies ratios are calculated and then multiplied: (1) the weighted daily breathing volume for animals vs. humans, (2) a deposition fraction ratio, (3) a normalising factor ratio, (4) an elimination rate ratio. HEC is derived by multiplying the exposure concentration from the animal study (c_T) by these four ratios according to the following formula:

$$HEC = c_T * \frac{AgV_T}{AgV_H} * \frac{NF_H}{NF_T} * \frac{ELR_H}{ELR_T} * \frac{DF_T}{DF_H}$$

where

T indicates animal data (“animal” in German language: “**T**ier”), H human data,

c_T is the exposure concentration from the animal study, for which we want to know the human equivalent,
AgV is the weighted breathing volume per day (German: "**g**ewichtetes **A**tem**v**olumen")
NF is a normalising factor,
ELR is the elimination rate, and
DF is the deposition fraction.

The most frequent starting points for interspecies extrapolation are rat studies. However, the use of mice studies is also briefly addressed.

The ratio for the weighted daily breathing volume (AgV_T/AgV_H)

In earlier versions of the HEC approach, the ratio for the weighted daily breathing volume has been a fixed value with data from one rat strain and from working persons. New data permit to consider more specific input from several rat strains and with different animal body weights.

For experimental data on breathing volumes of rats, a high variability is documented; various existing allometric regression formulae result in different breathing volumes at identical body weights. A currently suggested default value of 0.008 for this ratio apparently is a conservative approach. This value of 0.008 means that the weighted daily breathing volume in rats of 0.055 m³/day is assumed and compared to a weighted daily breathing volume of workers for chronic exposure of 6.57 m³/day (0.055/6.57 ≈ 0.008). Therefore, calculation procedures for various values are discussed considering the impact of strain and body weight on breathing volume of the experimental animals in the assessed study. Uncertainties of respective calculations are addressed accordingly.

The deposition fraction ratio (DF_T/DF_H**)**

The deposition fraction is calculated by dosimetric modelling in both species (e.g., rats and humans). This modelling includes mechanistic considerations and fluid dynamics with respect to sedimentation, impaction and diffusion of particles, with special consideration of the particle sizes and of the anatomy of the respiratory tract with different air flow characteristics in the upper, tracheobronchial and the pulmonary region.

Deposition within the HEC approach in this project is calculated by modelling with the "Multiple Pathway Deposition Model" (MPPD), which is freely available as updated version 3.04. Major changes compared to the former version (version 2.11) are described and the quantitative outcome is discussed for calculations of deposition for varying particle densities or particle sizes. Major areas of uncertainties are (1) inhomogeneity of deposition with potential "hot spots", which are not covered in subsequent HEC calculations, but might contribute to adverse effects and for which species differences are to be acknowledged and (2) hygroscopic growth of water

soluble particles, as this growth in particle size is not covered by MPPD calculations, but may significantly alter deposition patterns. Deposition in the lung is significantly influenced not only by particle size but also by particle density. In interspecies comparisons, it is relevant to know whether the relative fractional deposition ratio (rodents/humans) changes depending on size and/or density within the applicability range for default calculations.

The deposition fraction ratio as output from MPPD is a single ratio. However, in reality for both species (rodents and humans) there may be significant variability in the respective deposition fraction. Typical values for the deposition fraction ratio may be in the range of 0.2 to more than 1; this implies that the deposition fraction is frequently smaller in experimental animals than in humans, depending on the particle size in the experimental study. However, the term “more than 1” includes the possibility that a higher fraction of particles is deposited in the rodent lung than in the human lung.

From the limited data available for validation of the HEC calculations based on mice data and from an uncertainty analysis we conclude that interspecies particle deposition estimates based on mice data are associated with substantial uncertainty.

The normalising factor ratio (NF_H/NF_T)

Exposure needs to be quantified as a dose (measured in appropriate dose metrics) and needs to be related to a meaningful reference unit in the target organ (the lung). These steps are accomplished by assigning dose metrics to the deposited particles and by normalisation. Normalisation describes the reference unit for the deposited dose, for example, the alveolar lung surface area or the volume of the alveolar macrophages.

There is considerable variability in the data provided for either normalisation, leading to relevant uncertainty for this normalisation factor ratio. However, the most serious problem is to select the appropriate reference for normalisation. Choice of the appropriate reference might depend on the mode of action for the adverse lung effects, which is frequently insufficiently known. Further, if the average deposition in the respiratory tract is not determining the effect, but instead local deposition at hot spots is critical, this should be addressed by refined normalisation units.

Specifically, the influence of *particle solubility* is insufficiently correlated to the mode of action and to critical lung tissues. With solubility we refer to solubility in physiological lung fluids and not primarily to solubility in water. Solubility of the particle may greatly influence the mode of action in the respiratory tract (with respect to, e.g. primary target tissue, intracellular uptake, binding to proteins).

For poorly soluble and low toxicity particles (PSLT particles) the alveolar macrophage volume is frequently suggested for normalisation and the dose metrics used is corrected for particle density. However, many particles cannot be clearly identified as PSLT particles, and their toxicity mechanism is often unknown, although knowledge

of the chemical reactivity from either the surface of the particle or the solubilized particle is highly important for adequate HEC calculations.

The normalisation factor ratio is usually a quite large value as the much larger reference term for humans is divided by the respective smaller term in the experimental animal. For example, the alveolar surface area in humans may be estimated to be about 1,020,000 cm^2; the alveolar surface area in rats is said to be 4,000 cm^2. This results in a ratio of normalisation factors of about 250. Note, however, that this quantification is just one of many. If the total alveolar macrophage volume is used instead, most calculated normalisation factor ratios are even larger.

The elimination rate ratio (ELR_H/ELR_T)

The potential to exert adverse health effects in the lower respiratory tract will be greatly influenced by the residence time of the particles in critical regions of the lower respiratory tract. The *retention* of particles in the lung is directly correlated with the respective elimination kinetics. Therefore, the fourth step of the HEC calculation is the quantification of species differences in elimination rate.

Species differences in elimination rates were formerly mostly attributed to differences in mucociliary clearance, for which different half-lives of particles in the lung of rats or humans were observed. However, species differences in elimination from the lung may also result, e.g., from translocation to the interstitium or from different retention patterns due to binding of particles to biomolecules. If a compound is retained in the lung, but is quiescent due to binding to biomolecules, i.e. not biologically active, during certain periods of time, consideration of the retained dose would be misleading. Furthermore, the assumed first order kinetics may not always be justified, and the assumption of a multi-phase elimination process be more adequate.

PSLT particles are mainly eliminated via the mucociliary escalator. For this clearance mechanism species differences are well-known. However, species differences are less evident for other clearance mechanisms (e.g. translocation to the interstitium) and it is often assumed that there are no species differences for readily soluble particles or poorly soluble particles cleared partly by other mechanisms than the mucociliary escalator. However, this is regarded to be an oversimplification.

The elimination rate ratio for PSLT particles is usually reported to be about 0.15, which acknowledges the much longer retention of particles in the human lung compared to the rat lung. Again, there is considerable variability in this ratio. This quantification does not take into account species differences in translocation to the interstitium: in some cases, the particle fraction in the interstitium should not be regarded as eliminated from the lung but may contribute to adverse effects. The ratio of 0.15 might also not be applicable to soluble particles, but to date data are insufficient for a sound quantification.

Conclusions

With the HEC procedure the starting point is adapted: HEC aims at correcting the external concentration for differences between animals and humans with regard to the retained dose of particles in the lung it. However, because of the many uncertainties of the HEC approach (as shown in example calculations) we conclude that an improved starting point will not be easily established. Considering these uncertainties, the external exposure concentration in the animal study may be used as human equivalent concentration (HEC/concentration in animal study = 1). In a more conservative approach, a pragmatically derived assessment factor may possibly better reflect the overall uncertainties compared to highly uncertain, but scientifically refined quantitative ratios. Suggestions for such pragmatically derived assessment factors are given in this report. The consequences of either approach are presented and briefly discussed.

Abbreviations

AGS	Ausschuss für Gefahrstoffe (Committee on Hazardous Substances in Germany)
AGW	Arbeitsplatzgrenzwert (identical: German OEL)
AgV	Weighted breathing volume / d (gewichtetes Atemvolumen) / Tag
AgV_H	Weighted breathing volume for humans (H = humans)
AgV_T	Weighted breathing volume for animals (T = animals)
AM	Aleveolar macrophages
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
bpm	Breaths per minute
BW	Body weight
c_T	Concentration in animal study (T = animals)
COPD	Chronic obstructive pulmonary disease
DEF	Deposition enhancement factor
DF	Deposition fraction
DFG	Deutsche Forschungsgemeinschaft
DF_H	Deposition fraction for humans (H = humans)
DF_T	Deposition fraction for animals (T = animals)
ECHA	European Chemicals Agency

ELR	Elimination rate
ELR_H	Elimination Rate for humans (H = humans)
ELR_T	Elimination Rate for animals (T = animals)
EPA	Environmental Protection Agency (in the US)
GSD	Geometric standard deviation
HEC	Human Equivalent Concentration
HRTM	Human Respiratory Tract Model (developed by IRCP)
ICRP	International Commission on Radiological Protection
IIF	Interspecies interstitium factor
NEIR	interspecies normalisation and elimination rate ratio
NOAEC	No Observed Adverse Effect Concentration
LRT	Lower respiratory tract
MAK	Maximale Arbeitsplatzkonzentration (nonbinding OEL in Germany)
MMAD	Mass Median Aerodynamic Diameter
MoA	Mode of Action
MPPD	Multiple Pathway Deposition Model
MV	Breathing volume [m^3/d]
NF	Normalisation factor

NF_H	Normalisation factor for humans (H = humans)
NF_T	Normalisation factor for animals (T = animals)
NOAEC	No observed adverse effect concentration
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
OEHHA	Office of Environmental Health Hazard Assessment of the California Environmental Protection Agency
OEL	Occupational exposure limits
PAR	Proximal alveolar region
PMN	Polymorphonuclear neutrophils
PSLT	Poorly soluble and low toxicity (particle)
PSP	Poorly soluble particles
PU	Pulmonary region
TB	Tracheobronchial region
TCC	Total cell count
URT	Upper respiratory tract
VT	Tidal volume

1 Introduction

In the 1990ties, risk assessors developed a systematic procedure to derive a “human equivalent concentration“ (HEC), starting from effect concentrations for the lower respiratory tract determined in rodent inhalation studies. For regulatory risk assessment on inhaled particles, HEC calculations have already been suggested for the general population in 1994 (US EPA 1994). In 1999, based on deposition data of particles in experimental animals (e.g., Raabe et al. 1988) and humans (e.g., ICRP 1994) and respective airway and airflow modelling (Yeh 1980; Yeh and Schum 1980), and starting from provisional versions (Anjilvel and Asgharian 1995; Asgharian and Anjilvel 1998) the US Chemical Institute of Toxicology (CIIT) in cooperation with the National Institute for Public Health and the Environment from the Netherlands (RIVM) developed a Multiple Pathway Deposition Model (MPPD) (RIVM 1999). The HEC procedure included interspecies comparisons for a) deposition in the respiratory tract and b) retention, elimination, and clearance from the respiratory tract combined with several options for dose metrics (e.g. mass of particles or number of particles) and normalisation (e.g. to the lung surface area), all included in the MPPD software. However, it was also possible to limit the use of MPPD to deposition only (i.e. fraction of inhaled particles, which is deposited in a certain region of the respiratory tract) and supplement assumptions on elimination, dose metrics and normalisation from other sources as separate steps within the HEC calculation.

The HEC procedure including the MPPD deposition modelling has been used since 1999 mostly for specific areas of risk assessment of inhalation exposure to particles, with only few regulatory committees making use of this approach for standard setting: Systematic use of the HEC concept including MPPD dosimetry was established, e.g., by U.S. EPA for setting standards for the general population (US EPA 2004) and as part of the derivation procedure for occupational exposure limits (OEL) in Germany (FoBiG 2011). This German HEC approach has been presented in a guidance document on exposure risk relationship calculation for carcinogens in the lung (AGS 2013), but is not limited to carcinogens. The HEC approach was, for example, used to derive an OEL for “poorly soluble, low toxicity” particles (PSLT) in Germany (Hartwig 2012). However, this and subsequent applications in regulatory risk assessment induced some discussions on optimal parameter selection for MPPD modelling and on adequate procedures, e.g. for selecting dose metrics, normalisation and calculations of retained doses in the lung (e.g., Morfeld et al. 2015). Parts of the existing guidance on the German HEC procedure for workplaces (AGS 2013) were found to be not sufficiently elaborated to guarantee unambiguous application. Moreover, a more recent version of MPPD (version 3.04) was released in 2016 and needs inclusion into an updated handling strategy. Progress in inhalation toxicology and new information on biokinetics of particles are to be considered.

This report provides study results, discussion, information, and example calculations to develop an updated guidance for HEC calculations for particles in the lower respiratory tract (LRT) for occupational exposure scenarios. Recent data will be reported and critically assessed. In most instances, no final conclusion on a generally agreeable default procedure will be possible within the framework of this report, as

discussion on optimal use of the HEC approach and on its limitations is still ongoing; respective arguments and data will be presented and suggestions for handling will be provided, if regarded sufficiently qualified.

2 Definitions and Demarcation

2.1 Human Equivalent Concentration (HEC)

Within the context of this report the term “Human Equivalent Concentration (HEC)” will be used specifically to characterise the concentration of inhaled particles in the lower respiratory tract, where HEC is derived from rodent experimental data and biokinetic modelling according to the formula

$$HEC = c_T * \frac{AgV_T}{AgV_H} * \frac{NF_H}{NF_T} * \frac{ELR_H}{ELR_T} * \frac{DF_T}{DF_H}$$

where
T indicates animal data (“animal” in German language: “**T**ier”), H human data,
c_T is the exposure concentration from the animal study, for which we want to know the human equivalent,
AgV is the weighted breathing volume per day (German: “**g**ewichtetes **A**tem**v**olumen”)
NF is a normalising factor,
ELR is the elimination rate, and
DF is the deposition fraction.

This report will address all these ratios (AgV_T/AgV_H (Section 3), NF_H/NF_T (Section 5), ELR_H/ELR_T (Section 6), and DF_T/DF_H (Section 4)) separately in order to discuss procedures to quantify each of them (see Sections 3 to 6, for details). Finally, the aggregate procedure and the results for the ratio HEC/c_T will be discussed in more detail in Section 7.

Below, we will only provide a *standard* quantification procedure for HEC and for the single terms and ratios within this calculation: this will be called “*default* HEC calculation”. We will therefore define, e.g., particles properties and exposure conditions, for which default HEC can be calculated. For example, in this report a default HEC calculation is limited to a certain range of micro-sized particles only (Section 2.2). Therefore, if a HEC is to be quantified for nanoparticles, this may also be possible, but this is not regarded a *default* HEC calculation. Non-standard (non-default) interspecies extrapolation of human equivalent concentrations is not discussed in this report.

The procedure to calculate HEC in the way as described by the formula given above, we will call the **“4 ratios approach”** (product of 1. (AgV_T / AgV_H), 2. (NF_H / NF_T), 3. (ELT_H/ELR_T), 4. (DF_T/ DF_H)), in order to discriminate it from the “**aggregate 3 ratios approach**”, which is alternatively suggested in Section 7.4 for HEC-calculations.

2.2 Particle properties

The HEC *default* procedure as discussed in this report is limited to **particles sizes** with a mass median aerodynamic diameter (MMAD) or agglomeration diameter for nanoparticles of 0.5 - 2 µm (for justification see Section 4.8). The *more general* HEC procedure is linked to the **respirable particle fraction** and covers a broader range. The most recent MPPD software (version 3.04) provides means to calculate HEC within a particle size **range from 0.01 µm to 10 µm**, i.e. a considerably larger range than is covered by the default procedure. However, consequences for interspecies calculations for particle sizes beyond the mentioned smaller applicability range have to be discussed case-by-case and are not covered below. Specifically, for nanoparticle-specific transport and deposition, MPPD provides a separate adapted model for particles with a size of less than 0.1 µm, which is not addressed in this report.
Particles are assumed to be of spherical shape. Fibres are not covered. For **particles or agglomerates with irregular shape**, MPPD provides an "equivalent diameter model" and for **fibres** with an aspect ratio[1] larger than 3, different dosimetry assumptions are provided, but not addressed below.

The HEC default procedure is not limited to specified widths of the particle size distributions (defined by the **standard deviation** of MMAD), which means that it is not restricted to monodisperse or polydisperse particle distributions. However, no systematic testing of the uncertainties from wide distributions has been performed so far. It is assumed that studies with standard deviations of > 1.3 are not adequate for default HEC calculations, if the corresponding MMAD is close to the upper or lower applicability range (i.e. close to 0.5 µm or 2 µm).

Also, the HEC default procedure is not limited to a specific **bioaccessibility[2]** of particles in the respiratory tract. The procedure thus covers poorly and highly soluble compounds. However, specific uncertainties have to be addressed for highly water-soluble particles (see Sections 4.6, 5.5, 6.6, and 7.4). The HEC procedure, as discussed below, is linked to solid dry particles and does not address liquid aerosol exposure. The assessor should be aware of potentially differing solubility of particles in physiological lung fluids (i.e. epithelial lining fluid, interstitial fluid, lysosomal fluid), as those may influence mode of action, elimination and corresponding adequate normalisation (Sections 5.5, 6.6 and 7.4).

---

[1] **aspect ratio** of a geometric shape is the ratio of its sizes in different dimensions

[2] For the purpose of this discussion of respiratory effects, the term „bioaccessibility" is preferred to „**bioavailability**", as bioavailability usually refers to systemic (not to local) biological availability. However, some authors cited below use the term of bioavailability also for solubility in local respiratory fluids.

2.3 Effects in the upper respiratory tract

MPPD deposition calculations also report deposition in the upper respiratory tract (URT). Respective calculations point to significant species differences and should possibly be considered, if the critical respiratory effect of a particle is in the URT (e.g. Shang et al. (2015)). Apart from MPPD also other deposition models directly address the URT (e.g. Morris et al. 2010; Moss 2010). However, modelling of particle deposition in the URT is not further discussed below.

For gases and vapours with respiratory effects the URT often is the critical target. Asgharian et al. (2012) developed corresponding models, e.g. for formaldehyde, acrolein and acetaldehyde. These models also include dosimetry for the URT and the LRT region. However, this report is limited to particles. It is regarded worthwhile and relevant, to provide a concept for HEC calculations in the URT in future.

2.4 Nanoparticles

Specific conditions of nanoparticle HEC calculations are not addressed in this report. One reason for this is the applicability domain (particle size range) for this default approach (see Section 2.2). However, HEC calculations for agglomerates of nanoparticles with sizes above 0.5 µm are covered below.

It should be noted that workplace exposure usually includes only small fractions of single nanoparticles or agglomerates < 0.1 µm. There may be exemptions like welding fumes (Stebounova et al. 2018), for which a separate discussion is necessary (not addressed in this report). Numerous studies were performed with nanoparticles or agglomerates with smaller sizes than 0.1 µm; a significant part of those were studied under *in vitro* conditions. It is currently not suggested that such nanoparticles are to be handled as separate entity with significantly different properties from larger bulk particles (Gebel et al. 2014).

3 Weighted Breathing Volume: AgV_T/AgV_H

3.1 Breathing volume comparisons

Averaged weighted breathing volumes (AgV) (with the unit: air volume in m^3 per day) have to be multiplied with workplace particle air concentrations in order to determine the absolute amount of particles inhaled per day. The ratio AgV_T/AgV_H provides interspecies differences with respect to breathing volumes.

It should be noted that breathing volume also influences the deposition fraction in subsequent calculations (Section 4). Therefore, it is suggested that identical data and quantification procedures are applied a) to calculate AgV_T/AgV_H and b) as input to MPPD deposition modelling.

3.2 Current quantitative approach in Germany

In Germany, the HEC approach is currently used for interspecies extrapolation within the framework of deriving OELs for particulate substances (AGS 2013). Under chronic exposure conditions in a rat study with exposure for 6h/d the breathing volume is calculated as follows (AGS 2013):

AgV_T = tidal volume [mL/breath] x breathing frequency [breaths/min] x 60 min/h x 6h/day

= 2.1 mL x 102 1/min x 60 x 6h/d = 77 L/d = 0.077 m^3/d

with a default tidal volume of 2.1 mL/breath and a breathing frequency of 102 breaths/min. (0.214 L/min/rat). These values are from Long-Evans rats (Mauderly et al. 1979), but have been applied for any rat strain.

If exposure in the chronic rat study was at 5 days per week only, the average long-term AgV_T is calculated as follows (FoBiG 2011; Hartwig 2012):

AgV_T = 0.077 m^3/d x 5/7 = 0.055 m^3/d.

For chronic human exposure an average human breathing volume of 10 m^3/d is assumed, meant to stand for the breathing volume under light physical activity. This value is averaged over longer periods to calculate a yearly average AgV_H: (assuming exposure at 240 days per year):

AgV_H = 10 m^3/d *240 d/365 d = 6.57 m^3/d (FoBiG 2011; Hartwig 2012).

Therefore, in the current HEC default approach in Germany the ratio is set as follows:

AgV_T/AgV_H = (0.055 m^3/d)/(6.57 m^3/d) = 0.008.

The weekly exposure of the experimental animals may be 5 days per week, or, sometimes, 7 days per week. According to OECD 413 (90-day inhalation toxicity study) animals are typically exposed at 5 days per week. However, exposure at 7 days per week is also possible.[3] In contrast, OECD 452 (chronic toxicity testing) assumes exposure at 7 days per week, but would accept exposure at 5 days per week, if justification is provided.[4] If exposure was at 7 days per week the calculation of the AgV_T/AgV_H - ratio should be modified accordingly, i.e. 0.077 m³/d should be used.

Identical tidal volumes and breathing frequencies are used in MPPD, version 2.11: **for rats,** the default tidal volume is set to 2.1 mL and the breathing frequency to 102 breaths/minute. For 8 hours **human exposure** during workdays a breathing frequency of 20 breaths/minute and a tidal volume of 1040 mL is used as current default, resulting in a daily breathing volume of 20 x 60 x 8 x 1040 = 9,984,000 mL ≈10 m³ /day. For further discussion see Section 4.

There is a limitation of the current approach in Germany: Default breathing volume values are derived from only one rat strain (Long-Evans rats). Therefore, refinement is needed to cover breathing volumes for other rat strains (Section 3.3).

3.3 New data

No new relevant data on human breathing volumes (AgV_H) have been found in recent literature.

MPPD 3.04 permits the use of “Long-Evans rat” data for breathing frequency and tidal volume in experimental animals (rat). The data from Long-Evans rats discussed above (breathing frequency: 102/minute; tidal volume: 2.1 mL) are maintained and can also be used to calculate AgV_T. However, if specific body weights are provided and/or if other rat strains were used, MPPD 3.04 applies an allometric formula by Miller et al. (2014; 2013) to calculate breathing frequency and tidal volume for a given body weight. For example, for Sprague-Dawley rats (nose- or head-only-exposure) the following allometric formula is used in MPPD 3.04:

Tidal Volume (VT) [mL] = 1000 * (-0.060911+0.0013795*BW)/166 (Miller et al. 2014),

where BW is “body weight” in grams and 166 is a default value for breathing frequency of Sprague-Dawley rats, irrespective of body weight.

There are no clear rules provided in MPPD 3.04 how to calculate tidal volumes for other strains of rats but Sprague-Dawley or Long-Evans. Therefore, the Sprague-Dawley formula in combination with the specific body weight will be applied for any

[3] https://www.oecd-ilibrary.org/environment/test-no-413-subchronic-inhalation-toxicity-90-day-study_9789264070806-en

[4] https://www.oecd-ilibrary.org/environment/test-no-452-chronic-toxicity-studies_9789264071209-en

tested rat strain, if one uses the default automatic procedure of MPPD 3.04. See Section 4.3.2 for further discussion of the MPPD 3.04 calculation procedure and template. However, another allometric formula has recently been published by the Californian Office of Environmental Health Hazard Assessment (OEHHA 2018). The OEHHA regression formula has been derived from a large set of data from different rat strains including but not limited to male and female F344-rats, Wistar rats, Long-Evans rats and Sprague-Dawley rats. This regression results in an **Inhalation Rate (I)**:

$$I = 0.702 \times BW^{2/3} \text{ (unit: } m^3/day),$$

which could directly be used for AgV_T-calculation. This formula is linked to environmental exposure (24 h/d); for occupational exposure the value needs to be divided by 4 (if the study was performed with 6 hrs/day, which is the typical experimental design).

If the regression formula by OEHHA is used with a body weight of 250 grams a daily breathing volume of **0.28 m³/day** is provided (blue bold round mark, Figure 3-1). If the regression formula by Miller et al. is used (MPPD 3.04, Sprague Dawley Rat, 250 grams, asymmetric), those defaults lead to a daily breathing volume of **0.41 m³/day** (green bold round mark, Figure 3-1; VT = 1.71 mL; breathing frequency 166 * 60 * 6 = 59760 breaths/ 6 hrs; 1.71 x 59760/1000000 = 0.102 m³/ 6 hrs; breathing volume/24h = 0.102 x 4 = 0.41 m³/d). The original calculation in the German procedure and the default procedure in MPPD 2.11 (see Section 3.2) results in **0.31 m³/day** (red bold round mark, Figure 3-1; for 24 hrs. exposure, 0.214 x 60 x 6 h/d = 77 Liters/ d = 0.077 m³/ 6 h x 4 = 0.31 m³/d). This does not mean that either the breathing volume of 0.31 m³/day (current default) or the 0.41 m³/day (MPPD 3.04 for Long-Evans using the Miller formula) were incorrect, but it shows that the Miller et al. regression and the OEHHA-regression differ considerably. From this presentation it is also obvious that breathing volume calculations include high variability and high uncertainties. Similarly, MPPD 3.04 (ARA 2018, online - help-handbook) confirms that there is “considerable variability in published measurements of breathing frequency and tidal volumes”.

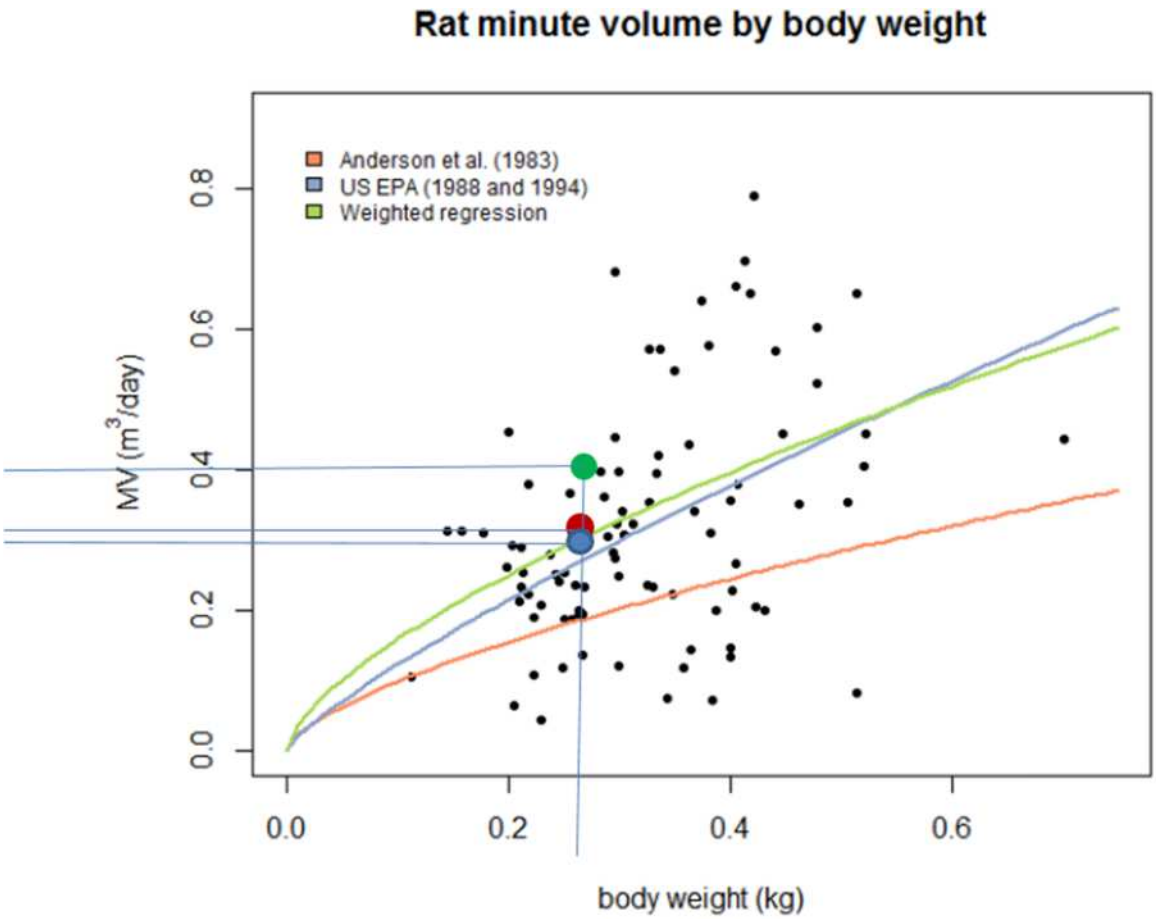


Figure 3-1 Inhalation rate per day (24h) as derived by OEHHA (2018) (green line); for interpretation of bold round marks (red, green, blue) (inserted roughly from graphical scale) see text, above. [note that MV (m^3/day) is described as breathing volume per day by the authors, not as breathing minute volume]; figure adopted with permission from OEHHA (2020; personal communication; April 8th, 2020; modified by inserted coloured bold round marks)

3.4 AgV_T / AgV_H from mice data

ECHA (2018) guidance provides default values for “inhalation volume/ hour” for (male) mice with 2.5 liters (default body weight: 30 grams). This corresponds to a breathing minute volume of 41.66 mL/min[5]. If breathing parameters are taken from MPPD 3.04 for a 30g mouse (BALB/c or B6C3F1), a default of 296.4 breaths per minute and a tidal volume of 1.799 mL are presented as default in the respective template for exposure. This tidal volume is regarded as incorrect.

The MPPD-manual for version 3.04 provides allometric formulae for breathing frequency and tidal volume:

[5] Table R.8-17;ECHA R.8 (2018)

- Breathing frequency (BF) [breaths/min]= 65.58 x $BW^{-0.4275}$, (BW in kg) and
- Tidal volume (VT) [mL/breath]=0.64175 x $BW^{0.29398}$, (BW in kg).

If BW= 0.03 kg are used for calculation, this results in a breathing frequency (BF) of 296 breaths per minute (confirming the default from MPPD template), but VT calculation results in a tidal volume of 0.229 mL, which is different from 1.799 mL (as documented in the MPPD 3.04 template). The product of BF x VT = 67 mL/min (breathing minute volume), which is only moderately different to the default from ECHA (2018), i.e., 41.66 mL/min (ECHA) vs. 67 mL/min (via allometric formula, as documented in MPPD 3.04-manual) (ARA 2018). Snipes (1989) provides a slightly lower breathing minute volume of 40 mL/min for mice.

The values in the MPPD 3.04 manual are also supported by Miller et al. (2016), who report 144-388 breaths per minute for mice (296 breaths are within this range) and a tidal volume of 0.218 mL for a 20 gram mouse (which is close to 0.229 mL from allometric calculation for the 30g mouse).

Hsieh et al. (1999) provides a different calculation formula to calculate a breathing volume (MV):

MV = 0.37 $BW^{1.36}$ [m^3/d] and Tidal volume = 0.0023 $BW^{1.36}$ (BW = body weight in grams).

For a 30g mouse, this results in a tidal volume of 0.235 mL (which, again, is very similar to the allometric calculation provided in the MPPD 3.04 manual). However, the breathing minute volume of 37.76 mL/min is quite low due to a low breathing frequency of 161 breaths per minute assumed in this calculation.

Kolanjiyil et al. (2019) report some data on breathing patterns for mice “at rest” or with “light exertion”. Mass flow rate is noted to be 66-125 mL/min, tidal volume between 0.2 and 0.22 mL and breathing frequency between 332 and 572 breaths per minute. As experimental animals are usually exposed at resting exposure conditions, the lower end of this range (i.e., 332 breaths per minute) can be used for further calculations. This is close to the default in MPPD 3.04.
An overview of the breathing volumes derived from above sources is provided in Table 3-1.

Note, that the breathing volume per day assumes 6 hours daily exposure, which is in agreement to the standard exposure duration in experimental animal studies.

Table 3-1 Daily breathing volume [m³/6h] for mice, different sources

Source	BW_{mouse} [g]	Breathing frequency per minute [bpm]	Tidal volume (VT) [mL]	Breathing vol. (m³/6h)	Remarks
ECHA (2018)	30			0.015	VT and bpm not reported
MPPD 3.04 calculation sheet	30	296.4	1.799	0.192	Values suggested for default. Calculated VT is questioned in this analysis. Different values can be entered in calculation sheet
MPPD 3.04 online manual	30	296	0.229	0.024	Different BW can be calculated (allometric formula)
Snipes (1989)	30			0.014	Calculated from minute volume of 0.04 L/min.
Miller et al. (2016)	20	144-388	0.218	0.011-0.03	Only range provided; note the BW of 20 grams
Hsieh et al. (1999)	30	161	0.235	0.0136	Different BW can be calculated (allometric formula)
Kolanjiyil et al. (2019)	?	332 (rest)-572 (light exertion)	0.2-0.22	0.024-0.045	

We suggest to use the MPPD 3.04 manual allometric formulae to calculate the daily breathing volume for mice. When the deposition factor is derived by applying MPPD 3.04 (Section 4.11), the default value in the exposure template needs to be corrected to the value as derived by the allometric equation, because the template default value is apparently too high. Defaults need to be agreed by further discussions and considerable variability in breathing rates and subsequent breathing volumes should be noted. If, for example, a breathing volume of 0.024 m³/day in the mouse is taken from Table 3-1, this should be transformed by a factor of 5/7 (average chronic daily exposure) to get AgV_T for mice: 0.024 m³/day x 5/7 = 0.017 m³/day (if the experimental chronic exposure is five days per week).

In Section 3.2, we derived the default breathing volume for human workplace exposure (6.57 m³/d; weighted daily exposure).

With a weighted breathing volume of 0.01714 m³/day in the mouse study, a weighted breathing volume interspecies ratio (AgV_T / AgV_H) for mice is calculated

AgV_T / AgV_H = 0.017 [m³/day] /6.57 [m³/day] = 0.0026

This default factor needs to be adapted to the specific body weight of the mice in the respective experimental study. This value is about threefold lower than the respective default value for rats.

3.5 Conclusions

From analysing updated breathing volume data in animals and, to a limited extent, in humans, we conclude:

- Human default assumptions on AgV_H have been maintained as in former assessments with a breathing frequency of 20/min and a tidal volume of 1040 mL, resulting in an inhalation breathing volume of 10 m³/d, which represents light physical activity. When averaged over chronic exposure periods (240 days per year) this results in a value of 6.57 m³/d. However, maintaining the well-established default of 10 m³/d, means relevant simplification. In reality, different breathing patterns (e.g., mouth vs. nose breathing) and individual differences in exercise and physiognomic parameters lead to a relevant range of AgV_H-values and, therefore, increase AgV_T/AgV_H-variability (not further assessed in this report).
- For experimental data on breathing volumes of rats, a high variability is documented; existing allometric regression formula result in different breathing volumes at identical body weights. The range of the ratio AgV_T/AgV_H is 0.0076-0.024. Therefore, the current default value of 0.008 for AgV_T/AgV_H apparently is a conservative approach.
- However, we propose to substitute the fixed value of 0.008 by a flexible value according to an allometric calculation.
- We suggest, <u>not</u> to switch to the OEHHA allometric breathing volume calculation because
    - The OEHHA-formula is less conservative for large body weights than the MPPD 3.04 formula
    - It is more complicated to use those OEHHA-derived values in combination with MPPD 3.04 software. The use of different breathing volume rates in MPPD 3.04 and for the standardized breathing value factor should be avoided.
- Therefore, the use of the flexible MPPD 3.04 generated values are proposed for AgV_T calculation, which can also be calculated manually by

  Breathing rate (166 breaths per minute) x 60 Minutes x 6 hours = 59670 breaths per working day

  Tidal Volume [m³] = (-0.060911+0.0013795*BW)/166000 (Miller et al., 2016),
- Deviations from this default calculation procedure should be considered,

- if explicit data are available for breathing frequency and/or tidal volume and/or breathing volumes are directly available from the experimental study,
- if exposure of experimental animals were not “nose- or head-only”, but “whole body”,
- if the human exposure scenario deviates from the default (i.e. 10 m^3/d breathing volume and exposure for 240/365d/year)

For AgV_T, in either of those non-default cases, the Help-Handbook-Online MPPD 3.04 provides supplemental allometric calculation formulae (note that those are linked to 24hrs exposure).

- If interspecies AgV_T/AgV_H is to be calculated from mice data, allometric formulae from MPPD 3.04 should be used for calculation of the animal breathing volume. The resulting interspecies factor for AgV_T/AgV_H is 0.0026 for a 30 gram - mouse to the human occupational scenario.

Table 3-2 Example calculations of weighted breathing volumes and AgV_T/AgV_H ratios for rat/human interspecies comparisons

Body weight (rat)	**AgV_T**	**AgV_H**	**AgV_T/ AgV_H**	**Comment**
Not assigned	0.055	6.57	0.008	Current default in Germany (AGS 2013; FoBiG 2011)
250 g	0.07	6.57	0.01	MPPD 3.04 (allometric formula by Miller et al.(2014))
250 g	0.05	6.57	0.0076	OEHHA (2018) (allometric formula by OEHHA)
500 g	0.08	6.57	0.012	500g (example default body weight for male rats, documented in ECHA (2018; Table R.8-17) MPPD 3.04 (allometric formula by Miller et al. (2014)
500 g	0.16	6.57	0.024	500g (example default body weight for male rats, documented in ECHA (2018; Table R.8-17) (allometric formula by OEHHA, 2018)

4 Deposition fraction (DF_T/ DF_H)

4.1 Deposition fraction – overview

Due to the specific anatomy and due to differences in air flow in the respective species, significant differences in deposited doses in the lower respiratory tract (LRT) exist between rodents and humans. Therefore, external exposure (corresponding to ambient air concentration) is regarded as a poor starting point for interspecies comparisons; the deposited dose in the pulmonary or total lower respiratory tract region may be more adequate. Therefore, the ratio of the deposition fractions is included in the HEC calculation procedure.

The term *fraction* within “deposition fraction” relates to the external (ambient air) exposure concentration (percent/100). However, in MPPD calculations the fraction can also be related to the “inhalable” particle concentration. This correction with respect to inhalability in calculations (“inhalability adjustment”) is discussed, when the influence of particle size is presented in a broader context (Section 4.9).

Currently, modelling of the deposited dose is an integrated element of MPPD. Some other modelling approaches are briefly mentioned in Section 4.2. For the German workplace HEC calculation procedure MPPD is applied and the calculation of deposition with this software is presented in more detail below (Section 4.3).

In guidances for interspecies comparisons with respect to particle effects in the LRT it is not precisely defined, whether deposition should be averaged for the total LRT or whether reference to local or regional area deposition within the LRT would be more adequate. This discussion is subdivided in 2 parts: discriminating the pulmonary region from the tracheobronchial region in Section 4.4, and discriminating average regional deposition from local “hot spots”, which leads to high inhomogeneity (Section 4.5).

Usually, solubility of particles is only discussed in the context of retention (because soluble particles are usually eliminated much faster from the lung). However, solubility of particles may also influence deposition patterns, as is documented in Section 4.6. Deposition in the lung is influenced not only by particle size but also by particle density. In interspecies comparisons, it is relevant to know whether the *relative* fractional deposition ratio (rodents/humans) changes depending on size (Section 4.8) and/or density (Section 4.7) within the applicability range for default calculations.

DF_T/ DF_H output from MPPD is a single ratio. However, in reality for both species (rodents and humans) there may be significant variability in the respective deposition fraction. This aspect is further discussed in Section 4.10.

Most interspecies extrapolations (and therefore HEC-calculations) are based on rat studies. However, MPPD also permits to calculate the deposition for mice (Section

4.11). Finally, conclusion from the discussed dimensions of deposition and the fractional deposition in rodents vs. men (DF_T/ DF_H) are presented in Section 4.12.

4.2 Alternative models for deposition modelling

Dosimetry modelling for the respiratory tract has been developed in the 1990th years, e.g. by the U.S. EPA, using the term “regional deposited dose ratio” (RDDR) with specific approaches for the upper and the lower respiratory tract (US EPA 1994). Within interspecies extrapolation, RDDR is used to adjust the animal deposited dose to a human-equivalent concentration (HEC). The RDDR software does not provide estimation of particle clearance or retention and the use of this approach has decreased over time (Kuempel et al. 2015).

Also, in the 1990th, the MPPD model has been developed (Anjilvel and Asgharian 1995; Asgharian and Anjilvel 1998; Asgharian et al. 2001; Price et al. 2002), which is discussed in more detail in Section 4.3. Overviews on similar modelling approaches including and in addition to MPPD are provided, e.g. by Isaacs et al. (2005), Kuempel et al. (2015), Fröhlich et al. (2016), and Lejon (2019).

For human exposures, the International Commission on Radiological Protection (ICRP) and the National Council on Radiation Protection and Measurements (NCRP) have, independently from each other, developed respiratory tract models for the use in radiation protection. However, these deposition data can also be applied for non-irradiant particles. Those models differ from the modelling within MPPD with respect to the mathematical model (NCRP/ICRP: semi-empirical; MPPD: deterministic) and lung geometry (NCRP/ICRP: symmetric lung geometry; MPPD: 5-lobe symmetric for default) (Asgharian 2018). The NCRP model (1997) gives rather similar results as the ICRP model (1994), but significant differences were found for nano-sized particles, where ICRP does not account for enhanced diffusional deposition (Yeh et al. 1996).

The ICRP-model is also described as Human Respiratory Tract Model **(**HRTM**)** and has been modified over the years (Gregoratto et al. 2010; Kuempel et al. 2001). Even though those models for human exposure are rather similar with respect to deposition, observations from workers still demonstrate deficits with respect to clearance (Kuempel et al. 2015) and therefore are currently updated (Bailey et al. 2007)[6]. These deficits are further discussed in Section 6.4 (retention and clearance).

All the models discussed above are validated mostly with poorly soluble particles. However, due to hygroscopic properties, deposition patterns may differ significantly for water-soluble particles. Winkler-Heil (2014) proposed deposition modelling specifically for hygroscopic particles, but, up to now, there is no user-friendly software available for routine application. The principles of a deposition model for hygroscopic particles have been described by Ferron et al. (2013)[7]. More details on the influence of solubility on deposition is provided in Section 4.6.

[6] No information on a realisation of this intended update is available

[7] Update information and a preliminary calculation sheet are available from

4.3 MPPD deposition modelling

4.3.1 MPPD Version 2.11 vs. Version 3.04

Applied Research Associates (ARA) issued an updated version 3.04 of the Multiple Path Particle Dosimetry Model in 2016, which is more closely described by Miller et al. (2016), Asgharian et al. (2014) and the online MPPD-Help-Handbook (ARA 2018)[8]. A number of changes and improvements compared to the former version 2.11 are included:

- Deposition modelling is now provided not only for the rat, but additionally also for B6C3F$_1$ and Balb/c mice, male rhesus monkeys, sheep, and pigs
- For the rat, deposition modelling is extended from Long-Evans rats only to Sprague-Dawley rats, where optional adjustments to different body weights can be considered and different modelling approaches (symmetric or asymmetric airway modelling) can be selected
- Further differentiated assessments for specific lung areas are possible, as more data on local alveolar surfaces are integrated
- Further, allometric calculation procedures are provided for, e.g., functional residual capacity (FRC) and upper respiratory tract deposition
- Modelling of deposition for toxic substances adhered to other particles is made possible (e.g. for environmental cadmium exposure associated with particulate matter)
- Deposition of particles with multimodal size distributions can be calculated
- Lymph node clearance is integrated as an elimination route
- The applicability domain of the model is increased to a particle size range of 0.001 µm to 100 µm
- Additional specific optional human exposure scenarios (like children's exposure profiles) can me modelled.

As MPPD is used for various scenarios, some of the recent changes may be very helpful, e.g., for site specific environmental risk assessments, but are most likely less important for interspecies extrapolation for regulatory standard setting. Note that MPPD is not always used for all steps of the HEC calculation. For example, in the German procedure, it was decided to assess interspecies differences in clearance not with MPPD (see Section 6.1).

4.3.2 MPPD (Version 3.04) application

No specific guidance on how to use MPPD (version 3.04) is included in this report. However, a few remarks on application of this software are useful to ensure identical results. In most cases, required input is identical to the earlier version (MPPD version 2.11). However, some modifications are summarized below:

---

https://www.helmholtz-muenchen.de/cma/forschung/topic-iii-aerosol-physik/projekte/index.html -> check for term: „Lung Deposition Model" and https://www1.helmholtz-muenchen.de/ioec/lung-deposition/hpldb06_i/index3i.php

[8] https://www.ara.com/mppd/

- For step: “Input data, airway morphology, rat” a specific choice is added in the field: “Model”, where “Asymmetric (lung model) for Sprague Dawley (rat)” can be selected. If agreed, the user is asked to provide a body weight in grams (output is erroneously given in kilograms). This will automatically change figures for FRC and URT due to allometric scaling (Please change number presentation from “decimal comma” to “decimal point”, as the program will turn to default, if you miss this correction).
- For step “Input data, inhalant properties, aerosol”, select “Inhalability adjustment” (yes), for either species (e.g. rat, humans) (for discussion see Section 4.9). Be sure to enter specific values for density and diameter (mostly given by MMAD). For closer discussion on the influence of density and particle size on deposition see Sections 4.7-4.8. For standard setting in default procedures do not change “aspect ratio”, single vs. multiple or multimodal, and do not mark for “equivalent diameter model”.
- For step “Input data, exposure scenario, constant exposure”, note that, for experimental animals, breathing frequency and tidal volume will be automatically adjusted to the body weight of the animal you have entered above (“Input data, airway morphology, rat”). But manual changes are permitted in non-default assessments. For humans, enter identical changes as requested in version 2.11 (e.g. for *workplace exposure scenario*: breathing frequency 20/minute and tidal volume 1040 mL, oronasal-normal-augmenter). Do not modify the standard inserted values for “acceleration of gravity”, “upright” body orientation, “inspiratory fraction” and “pause fraction”. For rodent input data, switch to “nose only exposure”, if applicable.

4.3.3 Quantitative changes (MPPD 3.04 vs. MPPD 2.11)

A quantitative comparison of the results between MPPD 2.11 and MPPD 3.04 has been performed. We analysed the influence of body weight and corresponding breathing volume/d on deposition fractions and on the ratio of the deposition fractions (DF_T/DF_H ratio) of the two versions. We compared results, when a) the allometric formula from either OEHHA (2018) for breathing volume was used or when b) the allometric formula from Miller et al. (2014) was applied. The latter approach is the default approach suggested in MPPD 3.04 (figures for tidal volume and breathing frequency automatically generated for a given body weight of rats). The former approach needs manual input of the tidal volume and the breathing frequency corresponding to the OEHHA calculation of the breathing volume/day (Section 3.3).

- Table 4-1 shows the output of MPPD 3.04 vs. MPPD 2.11 without consideration of the specific rat body weight.
- Table 4-2 shows the output of MPPD 3.04 vs. MPPD 2.11, where the specific rat body weight is considered (only possible in version MPPD 3.04) and the OEHHA calculation is used for the breathing volume.
- Table 4-3, finally, shows the output of MPPD 3.04 vs. MPPD 2.11, where the specific rat body weight is considered (only possible in version MPPD 3.04)

and the standard calculation suggested by MPPD is used for the breathing volume.

For all calculations deposition fractions in the pulmonary region (Alv) and/or in the tracheobronchial region (TB) are shown.

Without changes for body weight, for this example the results of the two versions were close to identical (Table 4-1).

However, if the specific body weight of F344 rats is taken into account (only applicable in MPPD 3.04, not in MPPD 2.11) and if breathing volume is calculated by OEHHA allometric regression, this results in relevant changes, in case only the pulmonary deposition or deposition in the TB region is assessed (Table 4-2). There is no relevant change, if deposition in the total LRT (TB+Alv) is considered (DFT/DFH: 0.45 in MPPD 3.04; 0.49 in MPPD 2.11). However, if only the pulmonary region is addressed, the difference increases (DFT/DFH: 0.32 in MPPD 3.04; 0.49 in MPPD 2.11). Note that even though the absolute deposition fraction is low in the TB area (1.66 or 2.82 percent, respectively), high DFT/DFH ratios can result if only the TB region is considered (0.47 or 0.8, respectively; Table 4-2). Similar ratios are observed, when the MPPD default breathing volume calculation is used instead of the OEHHA formula (Table 4-3).

Further, similar calculations indicate that the ratio DF_T/DF_H may be different up to about a factor of 2 (also for the pulmonary region only), depending on the input data, body weight and derived breathing volume and the specific region of the LRT addressed.

Table 4-1 MPPD comparative calculations (version 3.04 vs. 2.11).; input data: MMAD: 1.4 µm; GSD: 2.1; Density: 2.0 g/cm³; concentration: 0.067 mg/m³, default body weight assumptions accepted, parameters entered as requested in ERR-guidance (AGS 2013)

	MPPD 3.04			**MPPD 2.11**		
Depos. fraction	TB+Alv	TB	Alv	TB+Alv	TB	Alv
Human	0.1298	0.0354	0.0944	0.1300	0.0356	0.0945
Rat	0.0617	0.0162	0.0455	0.0631	0.0166	0.0465
DF_T/DF_H	0.48	0.46	0.48	0.49	0.47	0.49

Table 4-2 MPPD comparative calculations (version 3.04 vs. 2.11).; input data: MMAD: 1.4 µm; GSD: 2.1; Density: 2.0 g/cm³; concentration: 0.067 mg/m³, experimental body weight 434 grams F344-rat (NTP-study data; MPPD 3.04 only), breathing frequency: 140 (Mauderly et al. (1979) for F344-rats); Tidal volume as from OEHHA (2018) allometric formula associated with breathing volume at a given breathing frequency (see Section 3.3)

	MPPD 3.04			**MPPD 2.11**		
Depos. fraction	TB+Alv	TB	Alv	TB+Alv	TB	Alv
Human	0.1298	0.0354	0.0944	0.1300	0.0356	0.0945
Rat	0.0581	0.0282	0.0298	0.0631	0.0166	0.0465
DF_T/DF_H	0.45	0.80	0.32	0.49	0.47	0.49

Table 4-3 MPPD comparative calculations (version 3.04 vs. 2.11).; input data: MMAD: 1.4 µm; GSD: 2.1 ; Density: 2.0 g/cm³; concentration: 0.067 mg/m³, experimental body weight 434 grams F344-rat (NTP-study data; MPPD 3.04 only), breathing frequency: 166 (default SD-rats; MPPD 3.04); Tidal volume, default for body weight in MPPD 3.04 (see Section 3.3)

	MPPD 3.04			**MPPD 2.11**		
Depos. fraction	TB+Alv	TB	Alv	TB+Alv	TB	Alv
Human	0.1298	0.0354	0.0944	0.1300	0.0356	0.0945
Rat	0.06245	0.0315	0.0318	0.0631	0.0166	0.0465
DF_T/DF_H	0.48	0.89	0.34	0.49	0.47	0.49

An additional difference between MPPD versions was observed, when we analysed DF_T or DF_H values in relation to particle sizes (Section 4.8). Earlier versions of MPPD (i.e. versions 2.01 or 2.11) have shown minimum deposition at about 0.5 µm of diameter, with some increase in deposition with larger particle sizes and a local maximum at ≈ 2 µm (MPPD 2.11) or ≈ 3 µm (MPPD 2.01) (Figure 4-4). This local maximum was not observed with MPPD version 3.04, instead a monotonic decline of DF_T with increasing particle sizes occurs (
and Figure 4-3). For human data, MPPD 3.04 still demonstrates such a local maximum for low density particles (Figure 4-3), but not for high density particles (Figure 4-1 and Figure 4-3). Consequences of these differences are discussed in Section 4.12.

4.4 Deposition and region of the lower respiratory tract

As observed above (Table 4-2) and as supported by further MPPD calculations with other particle size distributions, deposition fraction in humans or rodents and DF_T/DF_H ratios can differ considerably in the various regions of the lower respiratory tract (LRT). There is currently no clear guideline to handle this uncertainty:

- For some respiratory effects, only deposition in the pulmonary region (PU) is relevant, for others the (lower) TB region should also be considered. The critical target cells are not always known, and more than one mode of action may be involved.
- For the same substance, different effects (e.g. carcinogenicity and COPD or inflammation) may occur at different sites within the LRT, but only one HEC is calculated.
- The most relevant site may differ between species, the PU region for the one species and the TB region could be more important for the other species (see also Section 4.5, below).
- Even though the absolute deposited dose in the one or other region may be low and does not considerably contribute to the overall particle load in the respiratory tract, this fraction may be decisive and may greatly differ between species, leading to changes in DF_T/DF_H ratios.
- Allocated deposition sites should fit to the subsequent steps for HEC calculations, i.e. normalisation and clearance. Therefore, the unit for normalisation (lung surface or lung plus TB surface or volume of macrophages etc.) should be selected in accordance with the critical deposition site. Clearance mechanisms will be different for particles deposited in the TB or in the PU region and more than one elimination pathway may be relevant and depend on the primary deposition site or the secondary site within the LRT.
- Subsequent calculations for retention and elimination adopt the original deposition fraction as a starting point. But this may not be correct after redistribution and translocation, where the fraction deposed in the TB or the PU region, respectively, may have changed with different relative body burdens.
- Specifically for larger particles (i.e. particle size > 2 µm), species differences increase: deposition in TB may be more important in rodents for the coarse particles and less relevant for humans and restriction to the pulmonary area may, thus, not be justified.

Currently, the default procedure in the German HEC calculations for deposition by MPPD includes the PU plus the TB region. This is discrepant from the respective assumptions used for normalisation and for clearance and, therefore, implies some uncertainties, which become more evident with the more recent MPPD 3.04 version (see example in Section 4.3.3). A more accurate region to refer to would probably be the PU plus the *lower* TB area. Specifically, for some tumours of the respiratory tract the region of concern may not include the upper TB region including trachea, but again, there are significant substance-specific differences and in most cases critical regions for deposition, translocation and final tissue interaction are not sufficiently

known. To *average* deposition fraction over all potential interaction sites may thus lead to unjustified species differences, and the HEC does not reflect the true (decisive) differences.

4.5 Inhomogeneous deposition

Selecting the adequate tissue or site in the LRT for calculating deposition is a key question. However, it is not limited to selecting the PU or TB tract as a whole but may need to be extended to critical spots within the PU or TB area, which may be crucial for adverse effects, whereas other areas are less relevant.

We have already raised this issue in an earlier discussion on HEC (FoBiG 2011), were we cited studies concluding that locally accumulated concentrations, e.g. at bifurcations of the respiratory tract, are more relevant than the overall average deposition level of particles within the LRT or PU region. Inhomogeneity of deposition can be expressed as “hot spot deposition enhancement factors” (DEF) with highly elevated deposition by 1 or 2 orders of magnitude. For example, Phalen et al. (2010) reported such DEFs >100. More recent documentations provide convincing evidence for such inhomogeneity, also demonstrating that the magnitude of those disparities depends on particle size, which makes it even more complicated to find an adequate DF_T/DF_H ratio for an appropriate local region of the respiratory tract (Dong et al. 2019). Balasházy et al. (2003) also reported high enhancement factors at hot spot areas and concluded: “Early histological studies …already indicate that neoplastic and preneoplastic regions predominate at bifurcation regions of the central airways“. Therefore, the hot spot concentrations in deposition may be crucial for subsequent effects. This is also supported by a recent study by Füri et al. (2020), who found serious inhomogeneity of radon deposition in the human lung: According to the authors „the study demonstrates that the cell nuclei receiving high doses are non-uniformly distributed within the bronchial airway generations. The results revealed that the maximum of the radiation burden is at the first few bronchial airway generations of the respiratory tract, where most of the lung carcinomas of former uranium miners were found.”

It was suggested to calculate deposition masses in close vicinity to the respective hot spots for HEC calculations. For example, Donaldson at al. (2008) proposed to use the “proximal alveolar regions” (PAR) for deposition normalisation (Section 5.4 and Section 6.7). However, those approaches have not yet been adopted in regulatory risk assessment guidelines.

4.6 Deposition and solubility

Currently, MPPD -deposition modelling in the lower respiratory tract of particles does not consider solubility. However, for water-soluble particles hygroscopic properties can influence deposition patterns (Varghese and Gangamma 2009). For example, water-soluble metal salts like cobalt chloride or zinc sulphate increase in size

("hygroscopic growth"), when entering the respiratory tract (Ferron et al. 2013). For sodium chloride with a dry diameter of ≈ 1 µm, a growth factor of 6 has been reported (Ferron et al. 2013). Winkler–Heil et al. (2014) described more specifically: "due to the variability and asymmetry of the human airway system, individual trajectories of inhaled particles are associated with *individual* growth factors, thereby enhancing the variability of the deposition patterns." For example, the authors described individual growth factors between 1 and 3.5 for particles with an initial dry size of 3 µm. Moreover, there are species differences in particle hygroscopic growth due to the different amount of time a particle travels through the regions with high relative humidity. The flow regime in the rat upper airways influences total and regional deposition much less than it does in human airways (Ferron et al. 2013). Because the turning points of the deposition probabilities differ between species, no linear relationship between hygroscopicity and the DF_T/DF_H factor can be established.

Therefore, hygroscopic growth leads to significant uncertainty and variability in HEC calculations for water-soluble particles. These uncertainties are currently not addressed in the HEC procedure and not covered in MPPD deposition calculations. Asgharian et al. (2014) explicitly confirms with respect to MPPD 3.04: "consideration was not given to the potential for differences between species of hygroscopic growth of particles, which could influence predictions of the respirable fraction."

4.7 Deposition and density

In Germany, the OEL for PSLT particles has been derived from animal data and HEC has been applied for interspecies extrapolation. This OEL is derived for a PSLT particle with standard density of 1 g/cm³ (and needs to be adapted for densities deviating from 1 g/cm³ by simply multiplying with the substance specific density). Therefore, we were interested to know about the influence of density on deposition in experimental animals and humans (influence of density will also be analysed within "dose metrics and normalisation"; see Section 5).

From a modelling approach by Braakhuis et al. (2014) there are indications that density at identical particle sizes significantly influences deposition. However, Morfeld et al. (2015) questioned such a significant impact. Therefore, we analysed the influence of density and MMAD for a broad range of densities, using MPPD version 3.04 (Table 4-4).

Table 4-4 Influence of density changes (0.1; 1; 5 g/cm³) in combination with particle size (MMAD: 0.5, 2, 3, 4 µm; GSD: 2) on deposition fractions and DF_T/DF_H. MPPD 3.04; body weight Sprague-Dawley rat: 370 grams (default allometric breathing volume]; 1 mg/m³

MMAD [µm]	GSD	Density [g/cm³]	DF_T	DF_H	DF_T/DF_H
0.5	2	0.1	0.0478	0.0511	0.94
0.5	2	1	0.0869	0.0827	1.05
0.5	2	5	0.1380	0.1316	1.05
2	2	0.1	0.0243	0.0781	0.31
2	2	1	0.0325	0.0866	0.38
2	2	5	0.0433	0.0993	0.44
3	2	0.1	0.0195	0.1237	0.16
3	2	1	0.0231	0.1294	0.18
3	2	5	0.0279	0.1378	0.20
4	2	0.1	0.001	0.0621	0.02
4	2	1	0.0115	0.0651	0.18
4	2	5	0.0135	0.0697	0.19

From this analysis we conclude:

- For small particles (MMAD ≈ 0.5 µm), depositions in rats and humans are similar, leading to DF_T/DF_H ratios close to 1 (i.e. 0.94 - 1.05) with only minor impact of the different densities.
- For particles with MMAD within the applicability domain of the HEC default procedure (Section 2.2), the influence of density on the deposition fraction ratio is limited, with an increase of DF_T/DF_H ratio by a maximum factor of 1.42 (for densities 0.1 – 5 g/cm³) or less for the examples calculated.
- For particle diameter > 3 µm, there is a larger influence of density on deposition in rats than in humans. At a density of 0.1, only a small fraction is predicted to be deposited in the rat lung (0.1%), whereas deposition is significantly higher in humans for such particles (i.e. > 6%).
- Generally, the influence of density on deposition is larger in the rat compared to humans.

These observations indicate that the impact of density on deposition is only small within the applicability domain of the default HEC procedure and is not linearly correlated with HEC. Therefore, the current handling of density within the PSLT OEL concept in Germany (normalisation of the OEL to a density of 1) is not justifiable by relative deposition, but, possibly, by the normalisation of dose with density (see Section 5.7 for further discussion).

4.8 Deposition and particle size

In Section 2.2 we defined an applicability range for default HEC calculations by particle diameters from 0.5 to 2 µm. This is suggested because of some substantial uncertainties for HEC calculations outside this range:

- Kuempel et al. (2015) describe different deposition mechanisms below 0.5 µm, with diffusion dominating at smaller sizes. The rapid increase of the deposition fraction at smaller particles sizes contribute to the overall uncertainty. The authors state: “The deposition of inhaled substances in the human respiratory tract depends on the aerodynamic diameter for particles larger than approximately 300–500 nm in diameter or on the diffusion diameter and density for smaller particles (including nanoparticles) …. The main deposition mechanisms are impaction, sedimentation, and interception for particles with aerodynamic diameters greater than approximately 500 nm, whereas diffusion is the predominant deposition mechanism for smaller particles…. These competing deposition mechanisms result in minimal deposition efficiency at approximately 500 nm”.
- Gregoratto (2010) states that translocation of particles to the lung interstitium is specifically relevant at particle sizes below 0.5 µm in humans and less relevant above. Therefore, default HEC application is specifically uncertain at this low end of microsized particles.
- OECD (2018) set a quality standard from animal inhalation studies with 2 µm MMAD as maximum particle diameter of exposure.
- Calculation results with MPPD version 3.04 predicts that deposition of particles > 3 µm in rats is minimal leading to extremely uncertain and high DF_T/DF_H ratios (Table 4-4).
- As indicated by calculations by ARA (the editors of MPPD 3.04) (Asgharian 2018, 2019), variability in individual deposition increases substantially in the range of very small (< 0.1 µm) or rather large (> 2 µm) respirable particles (Figure 4-5).

This limited size range for default extrapolations does not preclude case-by-case decisions to calculate HEC in the more uncertain range of particle sizes outside.

Interspecies comparisons on deposition of particles in the respiratory tract are clearly influenced by the particle size selected in the experimental animal study: if the animal (rat) study has used very fine particles (e.g. 0.1 µm MMAD), the DF_T/DF_H ratio is close to 1 (no substantial quantitative interspecies differences in deposition). If,

however, rats were exposed to more coarse particles (e.g. 3 µm MMAD), HEC will be much lower than the respective concentration in the animal study (i.e. $DF_T/DF_H << 1$). Table 4-5 and Table 4-6 and corresponding Figure 4-1 and Figure 4-3, respectively, demonstrate the relationship between MMAD and deposition for a broad range of particle sizes for rats and humans and two sets of data with different densities of particles and different rats strains (different rat body weight); Figure 4-2 demonstrates the corresponding changes in DF_T/DF_H for the data from Table 4-5.

Results shown in Figure 4-1 and Figure 4-3 were somewhat unexpected. All similar presentations (e.g., Greim 1997, Figure 4-3) of the deposition fraction in the pulmonary region for particle sizes above 1 µm MMAD showed a local maximum of deposition in rats at 2 or 3 µm diameter. This has also been demonstrated with using earlier versions of MPPD (2.01 or 2.11), as shown in Figure 4-4. No such local peak in deposition fraction has been found with the data from Table 4-5 and Table 4-6.

The specific background of this apparent change in deposition modelling in MPPD 3.04 has not been discussed in published comments.

Note that the particle size selected in the experimental animal study is regarded representative for all sizes of *respirable* particles; however, exposure to humans may be to larger or smaller *respirable* particles than those in the experimental study. Therefore the selection of particle sizes in the animal study will lead to more or less conservative OELs, depending on the particle size profiles of respirable aerosols at the workplace compared to the particle size profile in the “point of departure” animal study.

The effect of particle size on deposition should also be considered, when discussing the impact of *particle growth* for water-soluble particles on deposition (see Section 4.6): the resulting DF_T/DF_H ratio might be different for larger particles compared to the original (“dry”) particle size fraction. Therefore, it is hardly possible to make general predictions about quality and quantity of changes to the DF_T/DF_H ratio valid all over the range of respirable particle sizes.

Table 4-5 DF_T and DF_H and DF_T/DF_H ratios for a broad range from MMAD [rat: 0.1-3 µm; human: 0.1-5 µm) (Input data: rat body weight 400g; Sprague-Dawley; default assumptions on breathing volume, FRC and URT; default assumption for workplace scenario; particle density 4.0; GSD: 1.8; exposure conc. 1 mg/m³); PU-region and PU+TB-region

	RAT (DF)		**Hum (DF)**		**PU**	**PU+TB**
MMAD [µm]	PU	PU+TB	PU	PU+TB	DF(T)/DF(H)	DF(T)/DF(H)
0.1	0.2828	0.3645	0.298	0.4132	0.949	0. 882
0.3	0.1559	0.2072	0.1535	0.2181	1.016	0.950
0.5	0.1058	0.1593	0.1171	0.1689	0.904	0.943
0.8	0.0679	0.1224	0.1051	0.1494	0.646	0.819
1	0.0515	0.0983	0.1045	0.1459	0.493	0.674
1.3	0.0371	0.0769	0.1048	0.1431	0.354	0.537
1.5	0.031	0.0691	0.1047	0.1414	0.296	0.489
1.8	0.0233	0.0579	0.1032	0.1379	0.226	0.420
2	0.0186	0.0493	0.1013	0.1348	0.184	0.366
2.5	0.01	0.0288	0.0948	0.1255	0.105	0.229
3	0.00281	0.0169	0.0871	0.1154	0.032	0.146
3.5			0.0793	0.1056		
5			0.0579	0.0794		

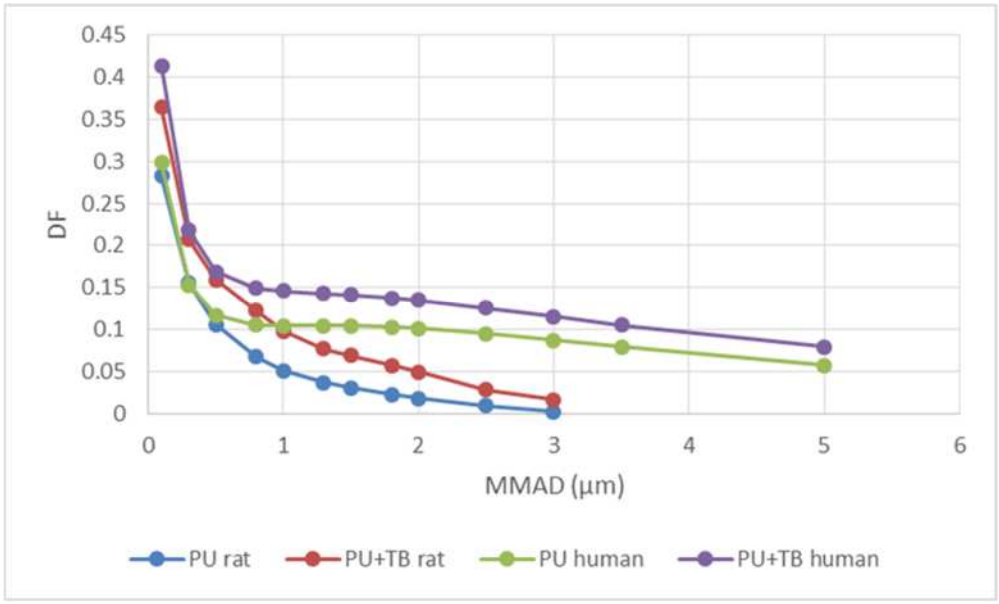


Figure 4-1 Particle Size and deposition in rat and humans from MPPD (3.04) calculations, density 4 (Input data: **Table 4-5**)

Table 4-6 DF_T and DF_H and DF_T/DF_H ratios for a broad range from MMAD [rat: 0.1-3.3 µm; human: 0.1-5 µm) (Input data: rat body weight default; asymmetric Long-Evans; default assumptions on breathing volume, FRC and URT; default assumption for workplace scenario; particle density 1.0; GSD: 1.5; exposure conc. 1 mg/m³), PU region only

	RAT	**Hum**	**PU**
MMAD[µm]	DF(T)-PU	DF(H)-PU	DF(T)/DF(H)
0.1	0.1878	0.1876	1.001
0.3	0.0901	0.0865	1.042
0.5	0.0683	0.0725	0.942
0.8	0.0588	0.0755	0.779
1	0.0554	0.0814	0.681
1.3	0.0501	0.0902	0.555
1.5	0.0451	0.0949	0.475
1.8	0.0374	0.0991	0.377
2	0.0322	0.1003	0.321
2.3	0.0278	0.0999	0.278
2.5	0.0244	0.0985	0.248
2.8	0.0194	0.095	0.204
3	0.0162	0.0921	0.176
3.3	0.0122	0.0871	0.140
5		0.0562	

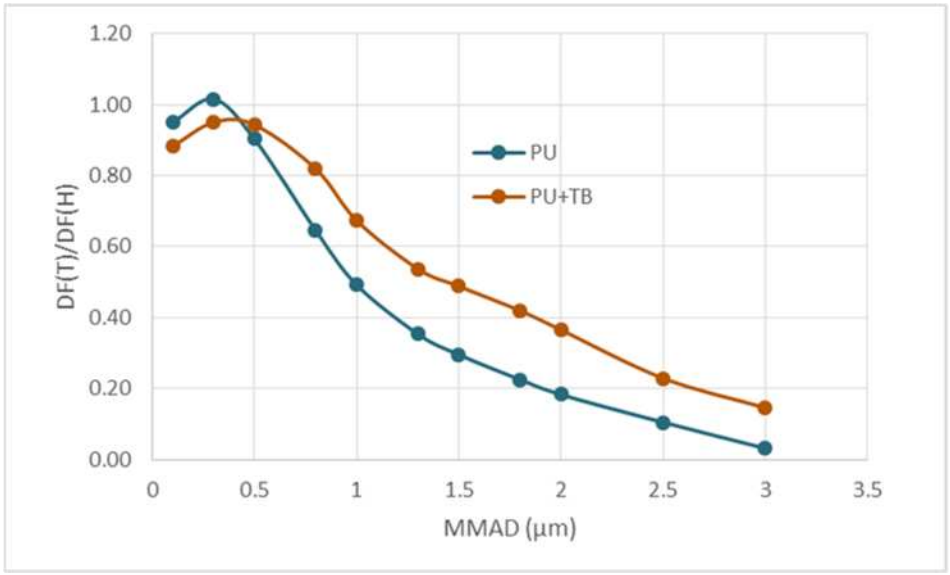


Figure 4-2 DF_T/DF_H ratios as a function of MMAD for the example data calculated in **Figure 4-1**, density 4 (Input data: **Table 4-5**)

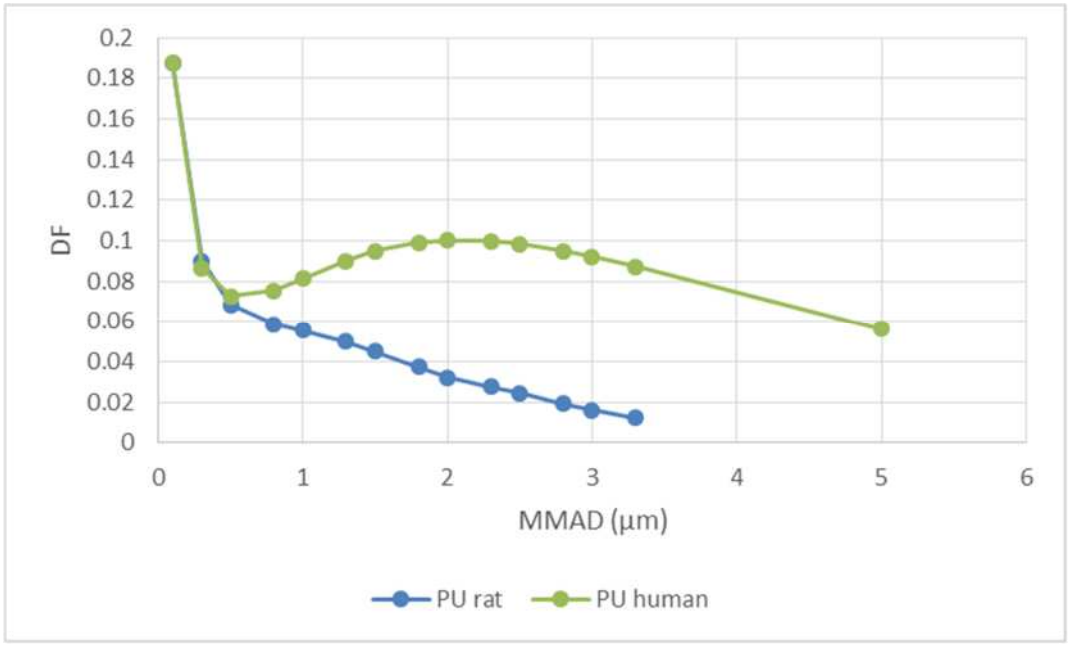


Figure 4-3 Particle Size and deposition in rat (DF_T) and humans (DF_H) from MPPD (3.04) calculations, density 1. (Input data: Table 4-6)

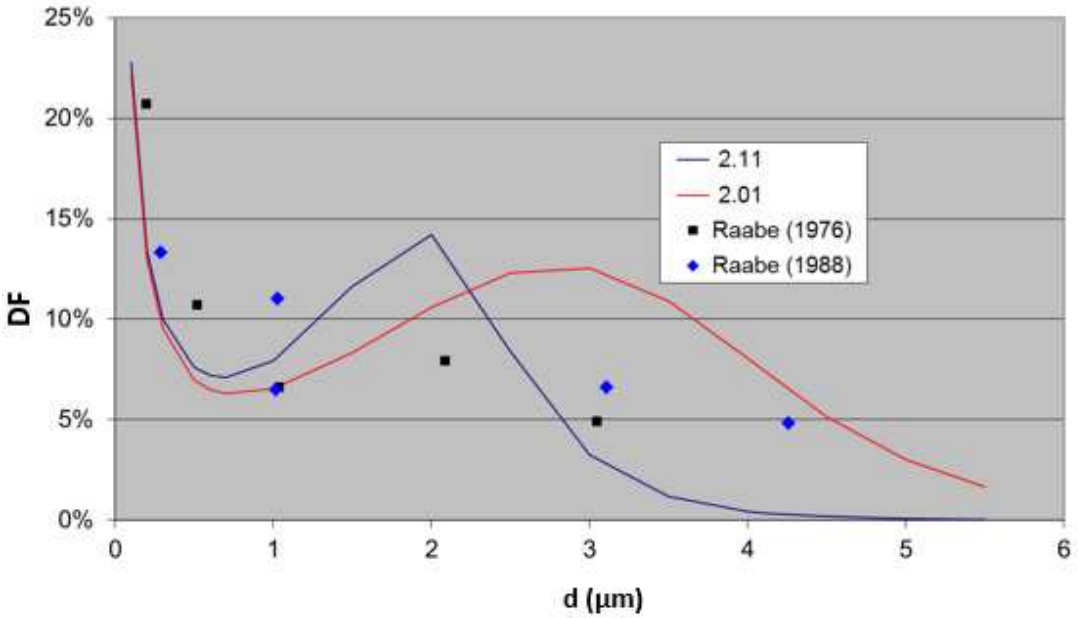


Figure 4-4 Deposition fraction in rats versus particle size according to MPPD version 2.11 or 2.01, resp.; both lines are non-monotonous with a second local maximum > 1 µm; exper. Data are from Raabe et al. (1988; 1976) (figure unpublished, sourced from internal discussions in German OEL-setting committee; d is aerodynamic equivalent diameter; discussed in Section 4.3.3).

4.9 Inhalability adjustment, applying MPPD

When applying MPPD 3.04 (or earlier versions) to calculate deposition the user can optionally tick “inhalability adjustment” or do calculations without inhalability adjustment. If calculations are performed without “inhalability adjustment” the inhalability fraction is set to 1.0, i.e. all particles are regarded as inhalable. If “inhalability adjustment” is ticked, the inhalable fraction is reduced, specifically in experimental animals, but only to a negligible degree in humans. If no “inhalability adjustment” is applied, the absolute and the fractional amount deposited in the respiratory tract of experimental animals is overestimated. Therefore omission of the “inhalability adjustment” has been criticized for certain applications of the MPPD modelling (Morfeld et al. 2015). However, the online-guidance to MPPD 3.04 does not generally request application of this inhalability adjustment: “this adjustment is relevant for particle sizes larger than 3-4 microns for rats and larger than about 8 microns for humans; the probability that particles larger than these are inhaled is less than 1.0 and decreases with increasing particles size…”.

In conclusion, there will be some difference in deposition calculations, if inhalability adjustment is ticked or not. However, the difference is rather small in the default region for HEC calculations (Section 2.2). If, however, MPPD is used beyond that

range, consequences will be significant. We suggest applying inhalability adjustment in any case, because this avoids underestimation of effects, even though the consequences will be marginal in the default region of applicability.

4.10 Deposition variability

The data presented above (Sections 4.3 to 4.8) demonstrate some uncertainty and variability in deposition calculations as part of the HEC procedure. Breathing volume and activity changes in experimental animals and humans, airflow modelling uncertainties, regional and local inhomogeneity in deposition patterns, influences of water-solubility on particle growth, influences of density and particle size, all contribute to a rather broad range of values for DF_T and DF_H fractions and consequently for the DF_T/DF_H ratio. Uncertainties may not be fully discriminated from variability.
Figure 4-5 provides an example for variability in the regional deposition fraction (intraspecies variability for different particle sizes and for PU vs. TB region). The author comments on the background for this variability: “explanation is that variation increases when external forces increase: diffusion for ultrafine, and impaction and sedimentation for coarse particles. These effects cancel each other out and deposition and also variation is reduced in the sub micrometer range.” (B. Asgharian; personal communication; April 8th, 2020). Such significant variability at the higher and lower range of particle sizes led to limiting the applicability range of the default HEC procedure, because selection of the most appropriate DF_T/DF_H ratio includes increased uncertainties (Section 4.4).

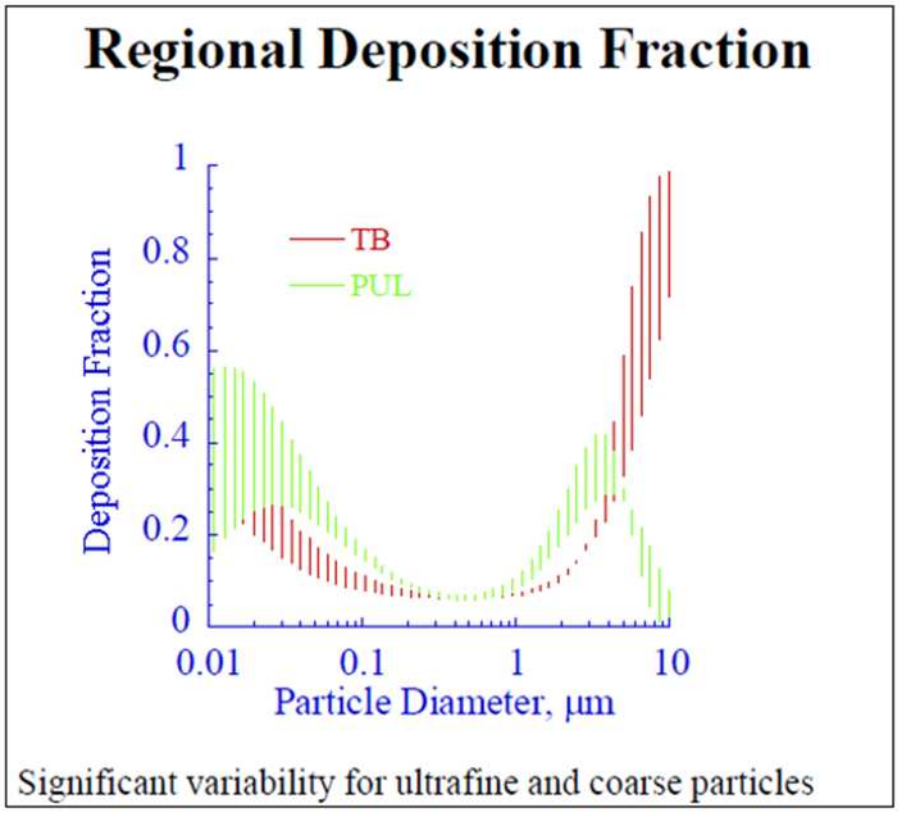


Figure 4-5 Human intraspecies variability in deposition, as indicated from MPPD modelling (Source: graphical presentation adopted from Asgharian (2019)); figure adopted with permission (B. Asgharian personal communication; April 8th, 2020)

Furthermore, differences in deposition due to compromised health of exposed persons at the workplace may contribute to overall variability. Generally, for OEL assessment scenarios, it is assumed that persons with moderately impaired respiratory health are frequently still attending daily work. Some respective comments are listed below:

- „Lung deposition may be altered in various pathological states, such as bronchitis, emphysema and fibrosis“ (Bos et al. 2019)
- "It is found that the PD [particle deposition] in models with… COPD has been disrupted by the geometrical changes and followed airflow alternations. …For COPD, the stenosis location determines the effects on DE [deposition efficiency] and DF [deposition fraction]. … DE increases with the particle size, and DE of the terminal bronchi is higher than that of central regions.“ (Zhang et al. 2018)
- "…These predefined parameters [e.g. in MPPD] do not include, for example, airway diameters and alveolar volume. …. This significantly limits the usefulness of these in silico lung models when moving from the healthy to the COPD lung." (Ganguly et al. 2019)

- “Uncertainties in the deposition of nanoparticles in the lung will remain due to considerable intersubject variability in lung morphology, breathing pattern, and possibly even circadian rhythms affecting the respiratory tract. This is particularly relevant for vulnerable subgroups of the population.” (Löndahl et al. 2014).

Variability in deposition is not covered by the HEC value and adds to the uncertainty of the HEC procedure.

4.11 MPPD deposition for mice

MPPD version 3.04 permits to calculate deposition rates for mice. Parameters were derived from BALB/c and B6C3F1 mice. The calculation procedure is largely identical to the procedure documented for rats (Section 4.3.2). However, it should be noted that the current template may contain an erroneous default value for the tidal volume of mice. We suggest to enter a default breathing frequency of 296 breaths per minute and a default tidal volume of 0.229 mL for a 30 g mouse in the exposure template of MPPD. With these data for mice and the particle characteristics also used for illustration in Section 4.3.3, Table 4-2, we calculate example DFT/DFH ratio as shown in Table 4-7. In this example, deposition fraction in the PU-region of mice is higher than in rats and deposition fraction in the TB-region is higher than in humans (and in rats). However, this relatively high deposition in the mouse TB region may be specific for only a small particle size range. Kolanjiyil et al.(2019) provide a general statement: “for micron-particles, the tracheobronchial deposition and alveolar deposition are significantly higher in the human lung than that in the mouse”. The authors also note: “the submicron deposition in the human distal lung airways is consistently lower than that in mouse-airway generations”. Mice data should not be used for species extrapolation at particle sizes above 3 µm. Similarly, Asgharian et al. (2014) give a warning: “there was little or no deposition of 3 µm and larger particles in the LRT [of mice]”. In a general statement, Kolanjiyil et al. (2019) acknowledge that only limited experimental data are available in the literature on mouse lung deposition. An analysis, comparing the deposition fraction from acute exposure of particles in the mouse lung from MPPD (version 3.0) and Raabe et al. (1988) by Ali et al. (2017) indicated major differences in deposition fractions in the two references. For example, MPPD calculated a PU deposition fraction of 12.37 % for particles with MMAD of 190 nm, whereas the Raabe et al. in vivo data indicated a PU deposition fraction of 45.4%. For a larger particle (MMAD: 767 nm) the MPPD calculation was 2.73 % (PU deposition), where Raabe et al. found 9.7%.

Table 4-7 MPPD calculations (version 3.04) for mice; input data: MMAD: 1.4 µm; GSD: 2.1; Density: 2.0 g/cm³; concentration: 0.067 mg/m³, body weight 30 grams mice, breathing frequency: 296; Tidal volume as from allometric formula (see Section 3.4)

	MPPD 3.04		
Depos. fraction	TB+Alv	TB	Alv
Human	0.1298	0.0354	0.0944
Mice	0.0789	0.0412	0.0377
DF_T/DF_H	0.608	1.164	0.4

4.12 Summary and conclusions on deposition

Deposition calculation provides important input for HEC interspecies extrapolation. However, several limitations, uncertainties and variabilities have to be acknowledged:

- In the default procedure, the particle size range is limited to MMADs of 0.5 - 2 µm, because
  a) at the low diameter end (< 0.5 µm), other deposition principles become predominant,
  b) at the low diameter end (< 0.1 - 0.5 µm), deposition fractions are highly influenced by the steep slope from maximum to local minimum deposition over an extremely small difference of size,
  c) below the lower diameter end, for the nanosized particles (< 0.1 µm) the deposition calculations are reported to be highly uncertain,
  d) at the high end of diameters, OECD guidelines request to limit diameters of test materials in animal studies to 2 µm,
  e) at the high end of diameters (> 3 µm) pulmonary deposition in experimental animals will be very low and may lead to overly conservative $DF_T/DF_{H\,r}$ ratios, and
  f) interindividual and intraindividual differences in deposition fractions due to breathing patterns and airway anatomy will rapidly increase beyond the default range of diameters in humans and therefore contribute to additional uncertainty of the DF_T/DF_H ratio.
- However, case-by-case calculations of DF_T/DF_H ratios with adapted parameters beyond the default HEC calculation range can be considered.
- The deposition modelling by MPPD does not include hygroscopic growth, which, however, is considered relevant, e.g. for water-soluble substances, as is demonstrated by significant changes of DF_T/DF_H, in response to minor particle diameters changes.
- Calculations with the MPPD 3.04 software disclose the significant influence of body weight (and, therefore, breathing frequency and tidal volume) on overall daily breathing volume in rats and, subsequently, on DF_T/DF_H. With the earlier

MPPD 2.11 it was not directly possible to take account of those strain-specific parameters.

- Calculations with MPPD demonstrate a decrease of deposition fraction at higher particle diameters in experimental animals in the pulmonary region, which is non-monotonous with a local maximum at 2 or 3 µm (MPPD 2.11, MPPD 2.01) and which is monotonous (no such local maximum) for MPPD 3.04. The background for this difference is unknown.
- Additional rules need to be developed: should deposition in the pulmonary region be addressed only, or should the tracheobronchial region (or parts of the TB-region) be included additionally.
- Moreover, local deposition at “hot spots” (e.g. bifurcations of the airways) may lead to highly inhomogeneous distributions, and average deposition in the lung or TB-region may be meaningless compared to such hot spot enhanced deposition sites. There are differences between species and mechanisms of such local depositions are not linearly correlated with average deposition. There are indications that (at least, some) neoplastic and non-neoplastic effects occur at such hot spots as target site.
- Breathing patterns, due to, e.g., nose or mouth breathing or exercise and influenced by physiological parameters, demonstrate high variability (human inter-individual differences) for submicron-sized and for large respirable particles. Impairment of respiratory health by particles will alter deposition fractions significantly, as, for example, has been shown for COPD.
- From the limited data available for validation of the HEC calculations based on mice data and from the uncertainty analysis by Kolanjiyil et al. (2019), we conclude that interspecies particle deposition estimates based on mice data are associated with substantial uncertainty.

5 Normalising Factor (NF_H / NF_T) and Dose Metrics

5.1 Normalisation and dose metrics – overview

Even after adjusting for species differences in *weighted breathing volume* (Section 3) and *deposition* (Section 4) the air concentration with unit mg/m^3 might not be the best measure to compare potencies of different particles in the respiratory tract. Exposure needs to be quantified as a dose (measured in appropriate dose metrics) and needs to be related to a meaningful reference unit in the target organ. These steps are accomplished by assigning *dose metrics* to the deposited particles and by *normalisation*. For HEC calculation, the interspecies ratio of the dose after normalisation are of interest, i.e. the ratio of normalising factors (NF_H/NF_T).

Note that in the current formula for HEC (Section 2.1), NF_H/NF_T does not include dose metrics explicitly and the ratio of normalisation factors calculated for a specific particle would be identical regardless of the dose metrics applied.

However,

- selection of specific dose metrics may be more or less appropriate for the various potential modes of action,
- selection of specific normalisation may be more or less appropriate for the various potential modes of action,
- the final HEC provided, e.g., as mass concentration (mg/m³) is different from the final HEC provided, e.g., as volume concentration (mL/m³) as dose metric,
- there are substance-specific differences in transformation of HECs provided as, e.g., volume concentration into, e.g., mass concentration,
- HECs for several particles can only be compared adequately, if provided in identical dose metrics.

Because similar considerations with respect to the mode of action are necessary for selecting dose metrics and normalisation, we discuss both steps of HEC quantification in this Section 5, even though transformation of HEC into the adequate dose metrics could also be analysed within the last step, the aggregated HEC calculation (Section 7).

Even though it would be helpful to generate results with *several* dose metrics and normalisations for one set of exposure data for discussion of possible modes of action (MoA), such complete data are rarely available and some have to be approximated, if needed. Moreover, regulatory purposes require a final output as mg/m³ (e.g. as an OEL, with *mass* concentration as final metric). Therefore, some suboptimal dose metrics may be considered to be acceptable considering easy calculations and easy transformation into pragmatic regulatory values.

The most serious problem in providing appropriate dose metrics and normalisation procedures is their dependency on a specific MoA and the typically large uncertainty about this MoA. If, for example, the impairment of alveolar macrophage function is the key mode of action, the appropriate normalisation may be the volume of alveolar macrophages, and the volume of the particles would probably be the adequate dose metric. But for many types of particles, the MoA is not unambiguously known and/or more than one MoA may be relevant.

Solubility of the particle may greatly influence the MoA in the respiratory tract (with respect to, e.g. primary target tissue, intracellular uptake, binding to proteins). In the discussion below, with solubility we refer to solubility in physiological lung fluids and not primarily to solubility in water. Therefore, suitable dose metrics and normalisation may have to be differentiated for the different solubility in lung fluids and different related MoAs. Further, if the average deposition in the respiratory tract is not determining the effect, but instead local deposition at hot spots is critical, this should be addressed by specific normalisation units. And if effect-related deposition (or translocation) is relevant also in the TB area, this calls for to a different normalisation compared to just the PU region reference.

We address dose metrics and normalisation separately and stepwise. This includes

- discussion of the German approach for an PSLT OEL with respect to dose metrics and normalisation (Section 5.2),
- discussion of various dose metrics and their suitability for different types of MoAs (Section 5.3),
- discussion of advantages and disadvantages of different normalisation units and ways to quantify them (Section 5.4),
- impact of solubility of particles on normalisation (Section 5.5),
- specific aspects on normalisation for HEC based on mice data (Section 5.6), and
- summary and conclusions (Section 5.7).

5.2 The German PSLT-approach and dose metrics/ normalisation

In 2014, in Germany an OEL for PSLT-particles was established, at

1.25 mg/m³ (respirable) PSLT particles for dust density of 2.5 g/cm³, i.e., 0.5 mg/m³ (respirable) PSLT particles for dust density of 1 g/cm³ (AGS 2014).

This regulatory OEL is only slightly different from the corresponding “MAK-Wert” by DFG (2019) of 0.3 mg/m³ (respirable) PSLT particles for a dust density of 1 g/cm³. Both values are based on a background paper, coedited by DFG and AGS authors (Hartwig 2012). This assessment was based on two approaches with similar results:

Approach A: based on two animal studies (with exposure to toner particles and titanium dioxide particles) animal NOAECs were reported and HEC was calculated

by the formula documented in Section 2.1 of this report. Dose metric was mass (mg) and normalisation was done by referring to the lung surface area. HECs were originally calculated as mass concentration in air (mg/m³). Subsequently, in a separate step, the results were transformed to a density of 1 (given density of toner dust was 1.2 and of titanium dioxide 4.3, respectively):

- The HEC for toner of 0.13 mg/m³ at density 1.2 was converted to a concentration of 0.11 mg/m³ for unit density.
- For titanium dioxide the original HEC of 1.1 mg/m³ resulted in 0.25 mg/m³ after transformation to density 1.

For justification, the authors explain: “Even if - in case of approach A - the deposited dose per square meter lung surface is calculated, for chronic exposure one needs to take account of the retained particle dose. The retained dose depends on the particle clearance. Particle clearance is influenced by the particle density / particle volume. Therefore density needs to be considered for approach A.” (non-literal analogous translation from Hartwig 2012)

Approach B: the approach was based on an analysis by Pauluhn (2011a) and postulated a MoA, under which the threshold for effects of PSLT would be linked to a particle volume of 6% of the total volume of alveolar macrophages, described as “overload-threshold” for inert particles by the authors. The threshold was calculated to be a volume-based generic mass concentration of 0.5 µl respirable particles/m³ × density. This results in a threshold of 0.5 mg/m³ for particles with a density of 1 in humans. Data of various PSLT particles including those for titanium dioxide (0.5 mg/m³ corresponding to 2.15 mg/m³ for a density of 4.3) fitted quite well to this postulated generic quantification. The result was also supported by data for nano-particle-agglomerates, if agglomerate density is used for dose metrics.

Approach B with dose metrics of particle volume and normalisation to the macrophage volume was applied in later German HEC calculations for nano-PSLT (AGS 2015). With respect to normalisation and dose metrics, we conclude that normalisation to **alveolar macrophage volume** and dose quantified as **particle volume** are the factual current default procedure of HEC calculations for PSLT particles in Germany, although the official default for PSLT and other particles in the generic HEC guidance still is mass for dose metrics and “alveolar plus tracheobronchial surface area” (Oberdörster 2010) for normalisation (AGS 2013).

5.3 Alternatives in dose metrics

MPPD, version 3.04, provides only *particle mass* related information as dose metrics of their output for subsequent calculations, i.e.

- Deposition fraction (mass deposited/ inhaled)
- Deposited mass (mg)
- Deposited mass rate (mg/min.)
- Deposited mass per surface area (mg/cm²)

- Deposited mass flux (mg/min/cm²).

Transformation to other dose metrics may be possible but are not included in MPPD 3.04. For example, number of particles or particle surface or particle volume are also frequently discussed as potential dose metrics. However, MPPD 3.04 also considers particle volume implicitly, as far as deposition calculations combine diameter (MMAD) and **density** for their calculations.

5.4 Alternatives in normalisation

As indicated above, the adequate unit for normalisation depends on the mode of action (MoA). Some dose metrics (Section 5.3) are closely correlated to normalisation. For example, particle volume is linked to the alveolar macrophages volume for normalisation to be meaningful, if macrophage particle loading is determining subsequent respiratory effects.

Generally, various alternative normalisation units are proposed, especially,

- Alveolar surface area or alveolar surface plus TB-area surface (m²)
- Lung weight (grams or kg)
- Lung volume (m³)
- Alveolar macrophages, number (n)
- Alveolar macrophages, volume (m³)
- Ventilatory units, number (n)
- Number of cells per lung (n)
- Surface area, type II epithelia cells (m²)
- Surface area, type I epithelia cells (m²)
- Proximal alveolar regions (PAR; m²) for hot spot correlation (Donaldson et al. 2008).

Many of those normalisation units are not regarded suitable for default calculations but can be considered case-by-case. The most frequently discussed units are i) Alveolar surface area or alveolar surface plus TB-area surface, ii) volume of alveolar macrophages, and iii) lung weight.

If those units were used for normalisation, numbers have to be selected to calculate NF_H/NF_T. Therefore, the current quantitative values are discussed below:

Lung surface area

There is no general rule whether the alveolar surface or the surface of the PU plus TB region should be used, if surface area is regarded the most appropriate unit for normalisation. In Germany, NF_H/NF_T = 150 was used for normalisation (default according to guidance; AGS 2013), which apparently has been chosen from PU plus TB region surfaces of rats and humans (Oberdörster 2010). Other measures and fractions are listed below (Table 5-1).

The difference between the minimum and the maximum is 140 vs. 349, indicating differences up to a factor of 2.5. Lung surface for humans has been discussed controversially (Bruch 2013; Gehr et al. 1978; Morfeld et al. 2015). Fröhlich et al. (2016) commented on the large values for human alveolar surface: “True alveolar surface available for gas is 20–50% smaller than the epithelial surface, depending on the level of air space inflation. At full inflation of 140 m^2, for instance, the “true” alveolar surface is only 70–100 m^2”. But still there apparently is no general agreement, as Morfeld et al. find that the large surface calculated by Gehr et al. in fact is not the epithelial surface but is the surface available for gas exchange.

Qualitatively, Oller and Oberdoerster (2010) suggest to select alveolar surface for normalisation: “The surface area normalized dose appears to be most useful for directly comparing doses in the respiratory tract between different animal species with vastly differing body sizes, like rat and human, considering that most effects are initiated by interaction of deposited particles with the epithelial cells of the respiratory tract and macrophages moving on the epithelial surfaces.” This concept is largely supported by Brown et al. (2005), stating: “If epithelial cells are the target, the TB- or alveolar surface area would be the most likely normalising parameter”. However, when the authors did not look to the *target*, but to the *cause* they suggested other normalising parameters for certain types of particles (below).

Table 5-1 Alveolar surface or PU- plus TB-surface in rats and in humans and NF_H/NF_T from different sources

Source	Humans [cm²]	Rat [cm²]	Ratio (NF_H/NF_T)	Remarks
(Hartwig 2012)	567780	2950	192	data from SD-Rats
MPPD 2.11 listed values	572220	2970	191	according to Fröhlich et al., (2016), referring to EPA (2004)
Oberdoerster (2010)	630200	4125	153	rat strain not identified in source, includes PU+TB area
Kuempel et al. (2015)	1020000	4000	255	also referenced in Fröhlich et al., (2016) (refers to Chen und Chen (2016)
EPA (2008)	540000	3400	159	no source or strain provided
Fröhlich et al., (2016)	700000	5000	140	Based on Lenfant et al. (2000): 1m²/ kg BW in mammals allometric scale
	780000			ICRP (Hum) according to Fröhlich et al. (2016), referring to Guha et al. (2014)
Morfeld et al. (2015)	1430000	4100	349	Morfeld et al., referring to Gehr et al.(1978) and Stone et al. (1992) as their sources

Volume of alveolar macrophages (AM)

AM volume is most frequently proposed for normalisation in HEC calculations for particles, but usually i) the analysis is restricted to PSLT particles and ii) the discussion is focusing on high exposure effects, where particle clearance via AM is impaired. The assumption then is that no adverse effects need to be considered at lower AM volume loadings for PSLT particles. The AM volume for normalisation consequently fits the assumed MoA (higher volume loading leading to persistent inflammation). However, AM volume is also more generally regarded relevant, as all particles (either poorly or readily soluble) are also taken up by macrophages. For example, some soluble particles may bind to endogenous proteins, therefore become less soluble and are subsequently phagocytosed by AM, contributing to AM load.

If AM volume is regarded the appropriate scale for normalisation, there is still some uncertainty about the quantification, due to different quantitative figures provided in different studies (Table 5-2). In Germany, recent HEC calculations for PSLT particles (nano-PSLT) have been performed using AM volume (Section 5.2). Specifically, the normalising factor by Pauluhn (2011b) of ≈ 1100 for NFH/NFT has been adopted. This estimate is based on AM volumes as reported by Krombach (1997), which

presented combined data from several sources. The authors derived the number of macrophages per lung by a body weight related regression equation for rats and from Oberdoerster (1995) for humans. Comparison of the ratio of normalising factors in Table 5-2 indicates a large range of values with a minimum of 278 and a maximum of 1110 (factor ≈ 4). This range demonstrates serious uncertainties about the most appropriate AM volume data to be used and generalized as a default, if AM volume is regarded to most appropriate parameter to normalise the deposited (or retained) dose.

Lung weight

Alternatively to lung surface or AM volume the lung weight has been proposed for normalisation (Brown et al. 2005). Beyond the parenchymal tissue the lung weight is influenced by the weight of the lung interstitium, which may also be the target of particle effects (Section 6.4). Kuempel et al. (2001) reported a lung weight of ≈ 1000 grams for humans and 0.9 grams for the rat, resulting in NF_H/NF_T - ratio of ≈ 1000. Similarly, Pott and Roller (2001) reported a human lung weight of 953 - 1200 grams and 0.9 - 1 gram for the rat (Wistar and Fischer), resulting in an identical normalisation factor.
It should be noted that this NF_H/NF_T of 1000 is at the upper end of the normalisation factor ratios derived from either lung surface or AM volume and therefore is not a precautious factor, even though it may be justified for some (not all) of the pulmonary effects from particles exposure.

Other units for normalisation

No further alternatives are discussed as “default normalisation units” here, as we found no applications of other normalisation units in practical particle OEL assessments. Some earlier suggestions and further data have been presented in former reviews (e.g., Brown et al. 2005; FoBiG 2011; Jarabek 1995; Kuempel et al. 2015).

Table 5-2 Alveolar macrophages volumes in rats and humans and NF_H/NF_T from different sources for quantification

Source	**humans ($x10^6$)§**	**Rat ($x10^6$)#**	**Ratio (NF_H/NF_T)**
Geiser, 2010 (F344-rats)*	1474 x 5990	639 x 29.1	310.23
Geiser, 2010 (SD-rats)*	1474 x 5990	1058 x 26.9	474.82
Miller, 2000 (SD-rats)	1474 x 5990	1161 x 26.9	278.4
Miller, 2000 (F344-rats)	1474 x 5990	882 x 29.1	344.01
Pauluhn,2011 (no specified rat strain)*	4990 x 7000	1166 x 27	1109.52
Kuempel et al., 2001 (no specified rat strain)**)	2500 x 7000	1000 x 26	673

*) cited from FoBiG (2011)
**) see also for further data
§) first term: Average AM volume [µm³]; second term : number of macrophages (x 10^6) /lung for a human with body weight of 70 kg
#) first term: average AM volume [µm³]; second term: number of macrophages (x 10^6) /lung for a rat with body weight of 370 grams

5.5 Influencing factors: mode of action and solubility

Solubility and bioaccessibility of particles in the respiratory tract are important factors in HEC calculations. As already discussed in Section 4.6, water solubility influences deposition. However, water solubility is a rather poor indicator to describe solubility in physiological lung media like epithelial lining fluid or in the intracellular environment from lysosomal fluid. Solubility in physiological lung fluids will also have significant impact on the mode of action in the respiratory tract and therefore on the most adequate *normalisation and dose metrics* and on *retention, clearance and elimination from the lung*.

In the current German HEC default approach, solubility is only addressed by its impact on the *elimination factor* and only based on *water* solubility (Section 6.3). The influences of solubility on MoA, normalisation and dose metrics are not discussed.

Consequences of solubility on normalisation, dose metrics and retention and the link to MoA will be further discussed in Sections 6.6 and 6.9. In Section 7.4, we propose a discriminating scheme to address MoA, normalisation and elimination for substances with different solubilities by integral categories.

5.6 Mice specific normalisation

Only few data for airway parameters are available for mice suitable for normalisation. Table 5-3 provides some data to compare the alveolar surface area and total alveolar macrophage volume in humans and in mice. However, the significant variability of the reported ratios of normalising factors (NF_H/NF_T) is to be emphasised, which extends if other human data but the data by Kuempel et al. (2001, 2015) are selected for human reference.

Table 5-3 Normalisation and ratio of normalising factors (NF_H/NF_T) in mice

Source	Airway parameter		unit	Ratio (NF_H/NF_T) *)	Strain/ Remarks
Asgharian et al. (2014)	alveolar surface area (∑ generation 15-21)	397.07	cm^2	2569	$B6C3F_1$-mice
		491.59		2075	BALB/c-mice
Hsieh et al. (1999)	alveolar surface area	1410		723	Not specified
Stone et al. (1992)	lung surface area	500		2040	Not specified
Knust et al. (2009)	total alveolar surface area	82.2		12409	CL 57 B6 mice [range: 63.3 cm^2, 101 cm^2]
Hsieh et al. (1999)	alveolar macrophages, total volume	1421 x 10^6	μm^3	12315	Not specified
Stone et al. (1992)		1430 x 10^6		12238	Not specified (19.2 grams)
Stone et al. (1992)		1300 x 10^6		13462	Not specified (two calculations in identical source)
*) assuming an alveolar surface area of 1,020,000 cm^2 (Table 5-1; Kuempel et al. (2015)) and a total volume of alveolar macrophages of 17,500,000 x 10^6 μm^3 (Table 5-2; Kuempel et al. (2001)) in humans					

5.7 Summary and conclusions on normalisation and dose metrics

Adequate normalisation and dose metrics selection are key steps within the HEC interspecies calculation procedure. However, as the most appropriate normalisation is closely linked to the mode of action and as mode of action often is insufficiently known or as more than one mode of action is relevant, the selection of the one or other normalisation unit and NF_H/NF_T ratio may often be premature or, at least, highly uncertain.

Only for unambiguous PSLT substances, remaining uncertainties are sufficiently limited to agree on AM volume for normalisation in default assessments.

Specifically, the influence of **particle solubility** is currently not sufficiently analysed on a generic level with potential consequences on

i) biokinetics in the lung,

ii) interactions with either alveolar macrophages or epithelium cells or both,

iii) endogenous protein interaction,

iv) direct vs. indirect starting points in activating immunologic response including macrophage- and PMN-activation,

v) the relevance of intracellular vs. extracellular contributions to effects,

vi) consequences (duration, pathways and species differences) in translocation and elimination from the lung (see Section 6.4), and

vii) gradual differences based on the rate of dissolution in the one or other lung fluid.

Considering that many particles cannot be clearly identified as PSLT particles, the influence of the chemical reactivity from either the surface of the particle or the solubilized particle is highly important for adequate HEC calculations. Normalisation to, e.g. the alveolar surface or AM volume or lung weight will shift the NF_H/NF_T ratio (see “aggregate 3 ratios approach” in Section 7.4).

6 Retention and Elimination (ELR_H /ELR_T)

6.1 Retention and elimination overview

The fourth step of the HEC calculation (after considering the *weighted breathing volume*, *deposition*, and *normalisation*) is the quantification of species differences in **elimination rate** (ELR_H/ELR_T). For a first order kinetic process, elimination rate is defined as $\ln 2/t_{1/2}$ (with $t_{1/2}$ = elimination half-life). Clearance half-life and elimination half-life are identical terms. The MPPD software includes a module to consider elimination. However, this step in HEC calculations may also be excluded and accounted for separately. This exclusion from MPPD is justified as the software does not specifically address the impact of different solubilities on elimination rate and there are concerns that species-specific clearance information is not updated even for PSLT particles (for further discussion: Sections 6.4 and 6.6).

The retained dose in the lung is regarded relevant, as it is assumed that adverse effects are linked to the long-term lung burden. However, this may not always be the case: if, for example, a compound is retained in the lung, but is quiescent, i.e. not biologically active, during certain periods of time, this reference to the retained dose would be misleading. Furthermore, the assumed first order kinetics may not always be justified, and the assumption of a multi-phase elimination process be more adequate. We will not address systemic effects from soluble particles after clearance from the lung, but there is a potential that particles are translocated to the lung interstitium; thus, the particles are cleared from the alveolar region, but interstitial effects still need to be considered. PSLT particles are mainly eliminated via the mucociliary escalator: for this clearance mechanism species differences in elimination half-life are well-known. However, species differences are less evident for other clearance mechanisms and it is often assumed that there are no species differences for readily soluble particles. However, these two categories (soluble particles without species differences and poorly soluble particles with fixed species differences) are regarded to be an overly simplified grouping. Any fixed elimination rate is an approximation only, with significant variability, e.g. due to individual airway anatomy and breathing pattern differences, inhomogeneous local retention in the various regions of the lung and due to changes in clearance due to illnesses. Consequently, we will discuss several topics in this Section on retention and elimination:

- Type of clearance mechanisms (Section 6.2)
- Current handling of the elimination rate in regulatory approaches (Section 6.3)
- Translocation to the interstitium and consequences for interspecies elimination rates (Section 6.4)
- Species differences in case of impaired clearance of PSLT substances (Section 6.5)
- Solubility and elimination rate (Section 6.6)
- Variation and inhomogeneity of the elimination rate (Section 6.7)
- Elimination rate in mice and interspecies extrapolation (Section 6.8)
- Summary and conclusions on retention and elimination (Section 6.9)

6.2 Clearance mechanisms and species differences

Because species differences in elimination rates were primarily discussed for PSLT particles, most discussions circle around mucociliary clearance. However, elimination and retention refer to various potential clearance mechanisms:

- Dissolution
- Physical translocation (e.g., mucociliary clearance)
- Phagocytosis by macrophages
- Lymphatic drainage (Jarabek 2016).

Usually, lymphatic drainage is combined with translocation to the interstitium to just one clearance pathway. However, this pathway should be subdivided and both steps should be separately addressed (Section 6.4).

As indicated, there are species differences in mucociliary clearance. However, quantitative species differences can also be expected for phagocytosis and translocation to the interstitium and to the draining hilar lymph-nodes (Nikula et al. 2001). Only for dissolution no species differences are known. As dissolution rarely is an isolated clearance mechanism, it cannot be generally concluded that there were no species differences for soluble particles (Section 6.6).

6.3 Current handling of elimination

There is no generic regulatory procedure for quantifying interspecies differences in elimination. In Germany, the following factors are used for calculating HECs. For PSLT particles, there are data for clearance rates in the rat without impaired clearance (often called “without overload”). This clearance (or elimination) half-life is provided with ≈ 60 days (AGS 2013).
In humans, clearance half-lives of 400 days or more are reported in literature (e.g., Hartwig 2012; Jarabek et al. 2005; Snipes 1989) and 400 days are used as default in the German procedure (AGS 2013). Clearance rate is calculated from elimination half-life by
Clearance rate (CL) = - ln (0.5) / elimination half-life (AGS 2013)

From this, species differences in clearance rates can be calculated:

$Clearance_T$= - ln (0.5)/60 d = 0.0116 per day
$Clearance_H$= - ln (0.5)/400 d = 0.00173 per day,
with $CL_H/CL_T = ELR_H/ELR_T$ = 0.00173/0.0116= 0.15

This ratio is used for poorly soluble particles at doses with no impaired clearance. There is no separate clearance interspecies factor for the “impaired clearance” situation. In the German guidance document on exposure-risk relationships it is explicitly stated: “in this case (i.e. at doses leading to overload situations) the identical factor is used as for unimpaired clearance” (Section 4.3 (3) in ERR-guidance; AGS 2013). However, effects observed in animal studies in the impaired clearance dose range are regarded as inadequate point of departure for quantitative cancer risk assessment. It is further stated that this elimination factor is highly

uncertain, if used for effect doses with impaired clearance. In addition, interspecies elimination rates ratio (ELR_H/ELR_T) are not changed in the current German HEC procedure, if the particle is poorly soluble but exerts its effects by some chemical reactivity. Again, this constant ELR_H/ELR_T is characterised by AGS (2013) as being uncertain, if used for such chemically active substances.
For highly soluble particles, ELR_H/ELR_T is set to 1. No species differences in elimination are assumed in this case. For particles with an "intermediate solubility" the default factor for poorly soluble particles is doubled, i.e. it is set to 0.15 x 2 = 0.3. This factor of 0.3 has been established pragmatically, with no specific empirical background. Solubility is used identically to "*water* solubility". However, there are no definitions for "soluble" or "intermediately soluble" provided. A case-by-case decision is suggested (for further discussion on the impact of solubility on the elimination factor, see Section 6.6).
Therefore, it should be acknowledged that the ELR_H/ELR_T ratio is rather uncertain for many types of particles. However, in the guidance document (AGS 2013) the value of 0.15 is regarded to be conservative, which may be questioned based on the observations with respect to elimination from the lung to the interstitium (Section 6.4).

6.4 Translocation to the interstitium and consequences for interspecies elimination rates

In an earlier report (FoBiG 2011), we have reported the data on human elimination half-lives for particles in some more detail. The default assumption of 400 days has been adopted from Hartwig (2012) derived from data. Other sources reported half-lives of about 30 days for 30% of the inhaled dose (phase 1 elimination) and of 700 days for 70% with a variability from 150 to 2500 days (phase 2 elimination; Snipes 1989). It had been acknowledged that some of this retained fraction may have been translocated to the interstitium, specifically in monkeys and in humans (Nikula et al. 1997). As translocation to the interstitium was not regarded as retention but as elimination from the lung, the estimated elimination half-life of 400 days was still regarded as conservative. Morfeld et al. (2015) criticized this 400 days value and suggested a clearance half-life of 250-300 days or 230 days, based on data from Gregoratto et al. (2010) and estimated from allometric considerations. In fact, Gregoratto et al. derived a clearance half-life of 300 days in humans, however, this was restricted to the fraction, which is cleared to the ciliated airways via the mucociliary escalator. In addition, the authors found a significantly longer elimination half-life (about 40% of the lung deposit) for insoluble particles "sequestered in the interstitium". For example, for radioactive ^{60}Cobalt-particles the total elimination half-life increased to about 1924 days, which is similar to the upper-range figure reported by Snipes (1989). The crucial question was whether the translocation to the interstitium should be assumed to be retention in a critical tissue of the lung or whether the translocated dose is to be regarded as already eliminated from the critical zone. This is directly linked to the question, whether effects can also occur in the interstitial region of the lung. If translocation to the interstitium is considered a way of elimination, then the ELR_H/ELR_T – ratio decreases (larger interspecies differences, because ratio is < 1), if just mucociliary clearance is covered, ELR_H/ELR_T

increases (less conservative). A potential highly significant difference in translocation to the interstitium in rats vs. humans is schematically shown in Table 6-1.
In more recent assessments, if was found that the amount translocated to the interstitium should not be regarded as eliminated (Gregoratto et al. 2011; Kuempel et al. 2001): specifically in humans it was observed that severe effects like fibrosis can be observed in the lung interstitium. Therefore, in the interstitial-sequestration model, Kuempel et al. (2001) included the interstitium, when calculating clearance half-time, whereas in earlier models (original HRTM-model and HRTM-modified model 1 from ICRP) this compartment was largely excluded ("Equivalent model"). Inclusion of the interstitium implies a more complex multi-phase elimination model instead of the simple first order kinetics, currently applied in the HEC approach (Kuempel et al. 2001)[9]. MPPD-software does not include this specific interstitial compartment (Fröhlich et al. 2016). Kuempel et al. (2015) found that "the estimates from these models differed by a factor of 2-3 with the interstitial-sequestration model predicting lower air borne concentrations associated with the working lifetime retained lung burden".

Fibrotic effects as observed in humans were not seen with PSLT substances in the rat at low concentrations or only to a marginal degree, whereas fibrotic effects in rats increased significantly only in doses well above the "overload threshold". From this it could be either concluded i) that the rat is not a suitable model to extrapolate interstitial effects to humans (this question is also raised by Bos et al.(2019)), or ii) that high "overdose" exposure in rats is needed to extrapolate such effects from rodents to humans, or iii) that an interspecies elimination ratio ELR_H/ELR_T well below 0.15 is needed to quantitatively include effects in the interstitium from rat to humans.

Nikula et al. (2001) showed different distributions within lung compartments for two PSLT substances (diesel soot (rats) and coal dust (humans)). The two PSLT substances were compared, as there were no interspecies comparative data for just one single compound available. For the parenchymal lumens in rats no influence on the percentage of retained PSLT depending on exposure concentration was observed (about 80% retention independently from particle concentration), whereas a significant decline of the retained particulate material (from about 40% to about 5%) was found for the human parenchymal lumens of the lung. The profiles for interstitium retentions were quite different: in rats, a low and nearly constant fraction of about 20% was retained by the experimental animals. In humans, the retained volume percentage increased from control to higher concentrations of the particulate material (from about 55% to about 90%). This comparison underlines the different relevance of the interstitium for retention of particles in the two species[10].

Bevan et al. (2018) analysed particle translocations of PSLT particles in rats and humans. They postulated highly different retention patterns of particles in the two species. The authors claim that this different distribution of the retained particles

[9] See Gregoratto et al. Gregoratto, D., Bailey, M.R., Marsh, J.W. (2010). Modelling particle retention in the alveolar–interstitial region of the human lungs. Journal of Radiological Protection 30, 491-512. for an illustration of the consequences of the different models on lung retention time (figure 3 in source).
[10] see figures 5 and 6 in Nikula et al. (2001)

leads to different toxicity potencies of PSLT particles in the two species[11]. It should be noted, however, that the illustration by Bevan et al. does not include concentration dependent changes in translocation and retention patterns.

Differently from PSLT particles, soluble metal particles like cobalt sulfate caused fibrotic effects in rats already at low concentrations (NTP 1998). However, for cobalt it is not clear, whether fibrosis is a relevant endpoint of respiratory effects in humans.

Byrne and Baugh (2008) and Dixon (2008) also describe specific evidence of translocation and fibrotic effects in experimental animals and humans in case of nanoparticles. However, also re-appearance of nanoparticles on lung epithelium from the interstitium has been observed and was probably mediated by macrophage-translocation (Geiser and Kreyling 2010; Riediker et al. 2019). Quantitative estimates on the relevance of this redistribution are not available.

Table 6-1 Schematic of rat and primate/human particle overload, as postulated by Bevan et al. (2018)

	Burden in rats	Burden in humans
Free particles in the alveolar space	+++	+
Particle-loaded macrophages	+++	+
Epithelial hyperplasia	+++	-
Interstitial transfer of particles	+	+++

6.5 Clearance impairments at high exposure concentrations

Clearance impairment is usually referring to impairment of mechanical clearance via mucociliary elimination. This impairment becomes obvious by a slower AM elimination rate. The slower AM elimination rate is not necessarily due to AM

[11] Bevan et al. (2018) mainly focus on the relevance of lung cancer as observed in the rat for human risk assessment. This report on HEC methodology does not discuss specifically certain endpoints like cancer and no potential species differences of such effects.

damage. In recent studies it is postulated that alveolar macrophages are not directly damaged or impaired in their phagocytic activity or increased in volume or less mobile at elevated lung particle loads (Li and Pauluhn 2018). These authors find that already a low lung particle load leads to increased influx of alveolar macrophages, polymorphonuclear neutrophils (PMNs), and cytokines in the lung (summarized by an increase in “total cell count”, TCC), which, in turn, leads to reductions in mucociliary clearance just because of the increased number of cells to be eliminated per unit of time. At high TCC levels, where adaptive responses are not sufficiently protective anymore, there will be inflammatory injury of the lung. Li and Pauluhn (2018) specifically focus to the relevance of PMNs: “neutrophils are primary perpetrators of inflammatory injury in the lung”. An increase of about 4% of PMNs may be indicative for some first adverse effects in the lung. Similarly, an increase of TCC by 6% in the lung may also be indicative for this particle effect, where clearance half-life also increases significantly. The increase in TTC by 6% coincides with the 6% volumetric load of the alveolar macrophages (“displacement volume”). This observation therefore maintains the 6% AM volumetric load from Morrow (1992), but provides only a coincidence by chance with this early figure, which is not mechanistically linked to an impairment of the alveolar macrophages (Section 5.3). This new interpretation of the “overload effect” gives rise to some doubts, i) about the mechanistic interpretation of the “overload” effect, ii) about the justification of the particle volume being the most appropriate dose metric, iii) about the AM volume as being the best unit for normalisation, as it is not clearly mechanistically linked to the observed effects, and iv) about impaired mucociliary clearance being the earliest and most significant adverse effect from particle exposure (Section 5.3), while inflammation of lung epithelium may occur at similar concentrations and is more clearly regarded as being an adverse effect. However, other authors have postulated damage of alveolar macrophages from particles and assume that clearance hindrance is not primarily a secondary effect from the increase in TCC (Bos et al. 2019).

It should be noted that an increase in elimination half-life in the rat is not always an indication of a mere particle effect. Particles with some chemical reactivity (from solubility or just particle surface reactivity with biological matrices) can lead to direct damage of the alveolar macrophages and, thus, to impaired clearance. Therefore, it may be premature to assign all clearance impairments at elevated exposure levels to a PSLT effect.

6.6 Elimination of soluble particles

In MPPD 2.11 consideration of clearance focussed on poorly soluble particles and could not be changed by the user (Miller et al. 2016). Therefore, MPPD use was limited to PSLT substances or to model deposition. However, in principle, version 3.04 of MPPD includes a formula to calculate clearance, which allows the user to specify values for certain constants as well as for the mucous velocities for the TB region, thereby extending the clearance modelling of MPPD to any type of particle (Miller et al. 2016). However, such “specific constants” to calculate clearance are usually not available, not specified for the two species, and probably would differ for (a) water soluble particles, (b) particles that are soluble in the alveolar lining fluid, (c)

particles that are soluble in the lysosomal fluids, (d) particles that are soluble in the interstitium in combination with subsequent reactions with biological matrices, which may further modify clearance velocity. Therefore, it is concluded that MPPD is still not a suitable tool to calculate clearance for the different types of particles with varying bioaccessibility in different regions of the respiratory tract.
As indicated in Section 6.3, the consequences of solubility on clearance time and on the most relevant mode of action are not well known, specifically at intermediate solubility in physiological fluids. Water solubility is a poor indicator of solubility in the various compartments of the respiratory tract. It is generally assumed that there are no major species differences in elimination half-life from the lung for highly soluble particles, if dissolution determines clearance time (Oller and Oberdörster 2016). However, even for highly soluble particles, dissolution is not the only determinant of retention and elimination rate, with variable consequences for interspecies differences:

- Cadmium oxide, which is poorly soluble in water, is eliminated in animal species at a similar rate as is the highly soluble cadmium chloride (Oberdörster 1988).
- The highly soluble cadmium chloride is eliminated much faster in rats compared to dogs or monkeys (Oberdörster 1988).
- Tricobalt tetraoxide, which is moderately soluble in water (1.6 mg/L in water, 20°C) is eliminated much faster in rats, hamsters and mice compared to dogs, guinea pigs, baboons, and men (Bailey et al. 1989).
- Particle size may greatly influence solubility: as shown for tricobalt tetraoxide in vitro, intracellular solubility resulting in substantially different dissolution rates after 2 weeks, when 50% vs. 5% vs. 3% vs. 2% of the original particle mass for 0.3 or 0.7 or 0.8 or 1.7 µm-particles, respectively, were solubilised (Kreyling et al. 1990)
- Even highly soluble particles may be translocated and retained in the interstitium: the highly soluble cobalt sulphate is predominantly eliminated to the interstitium in rats resulting in long-term effects in this tissue, but similar elimination kinetics in humans are only observed for insoluble cobalt compounds or tungsten carbide alloys (NTP 1998, 2014), with insufficient human data for cobalt sulphate
- Chemically active soluble substances may damage alveolar epithelial cells, which leads to major changes in elimination kinetics. Theoretically, the alveolar epithelium could be damaged to a degree that solute clearance becomes limited by the endothelial barrier of the pulmonary capillaries (Oberdörster 1988)
- In impaired lung tissue (e.g. from smokers) elimination of soluble particles is increased due to effects on the alveolar epithelial barrier of mediators released from activated AM and solid particle clearance decreased, due to impaired AM function (Oberdörster 1988)
- Adsorption of soluble particles to other particles (e.g. benzo(a)pyrene to diesel particles) or to endogenous biomolecules (like metallothionein, enzymes like phosphatase) may significantly alter elimination rate time of particles and MoA

of pulmonary effects (Galle et al. 1992; Oberdörster 1988), e.g., for metal compounds (Beyersmann and Hartwig 2008).

- Nanoparticles with different solubility properties were analysed for their elimination kinetics. For example, slow dissolution (abiotic dissolution ≪30% per 7 days, or even no apparent dissolution) of barium sulphate and silica dioxide was followed by re-precipitation and transformation. In contrast, e.g., zinc oxide or copper oxide showed high dissolution and clearance (abiotic dissolution ranging from 30% to 100% after 7 days) Particle size had a relevant influence on dissolution properties (Koltermann-Jülly et al. 2018).

As mucociliary clearance by alveolar macrophages contributes to elimination of moderately soluble particles and as also readily soluble particles may be eliminated in parts via AM transport, if those reacted with endogenous proteins, there is probably only a small portion of particles where no species differences are expected. However, there are only few adequate lung clearance data available for interspecies comparisons on clearance and conclusions are, therefore, highly uncertain.

6.7 Variation in clearance due to respiratory illnesses, individual differences and local inhomogeneity

Average elimination data and average interspecies elimination rate ratios (as provided by ELR_H/ELR_T) may not be representative and meaningful, if, in truth, local elimination at certain hot spots determines respiratory effect potency (Sections 4.5 and 5.1). For hot spots and PSLT-particle elimination Donaldson et al. (2008) suggests the **alveolar region proximal to those hot spots** (PAR) to be more adequate to calculate meaningful species differences: "*The proximal alveolar region (PAR) of the lung has been identified as a key site for the retention of respirable particles, as it receives high deposition but has slow clearance compared to the larger airways*" (Donaldson et al. 2008).

In addition, it should be noted that there is individual variability in elimination due to personal airway anatomy, breathing patterns and potential airway impairments from respiratory illnesses. "ICRP recommends reducing the clearance rate by a factor of two when estimating the retained particle dose among individuals with COPD" (Kuempel et al. 2015). There exists no adequate aggregate information about the variability in clearance rate in experimental animals and the consequences of this variability on ELR_H/ELR_T – ratio variability. Within the framework of this study, no substance specific data on clearance variability in rats were retrieved and statistically analysed.

6.8 Interspecies differences in elimination rate from mice experimental studies

Elimination half-life data from the lung of mice for PSLT particles are reported in literature (Benson et al. 1995; Snipes 1989; Snipes et al. 1989). However, there are only limited data available, which may be less representative than those for the rat.

Examples for very similar elimination rates for PSLT particles as compared to rats are provided in literature (Snipes 1989). Changes of half-life due to solubility have to be assessed case-by-case. There is more evidence for mice than for rats that particles are translocated to the interstitium.

For the purpose to calculate an example HEC in this report (mice to human – extrapolation), we assume an identical elimination half-life for mice and rats (Section 7.8.4).

6.9 Summary and conclusions on retention and elimination

The current German HEC procedure (AGS 2013) to calculate ELR_H/ELR_T for PSLT substances with a default rate of 0.15 is apparently quite conservative, if elimination to the interstitium is not considered. In this case, the factor could be increased to 0.2 (=60/300), due to shorter clearance time in humans. However, because of significant species differences in translocation to the interstitium and potential interspecies differences in adverse effect potency in the interstitium, these values for ELR are highly uncertain, even for PSLT particles and also at lower concentrations (below concentrations that lead to impaired AM clearance). Therefore, in addition to a default ELR_H/ELR_T of 0.2 for AM clearance, the probability of relevant effects of PSLT particles in the interstitium should be considered in both species, which may increase interspecies differences.

For particles with low solubility, which are, however, chemically reactive in the alveolar region, there are insufficient data to conclude on a default factor. If clearance time in the rodent is increased (i.e. significantly higher than 60-90 days), this may be due to general high dose particle effects associated with impaired AM clearance and due to increased TCC in the alveolar region (“overload effects”), but it may also be due to chemically induced damage of macrophages or epithelial cells with subsequent reduced elimination. It is often not known whether identical reductions in elimination rate takes place in rats and in humans.

Similarly, for substances with intermediate solubility in physiological fluids or intermediate water solubility the reduction in species differences is not well known, but it is obvious that the mode of action may be different compared to PSLT particles. Probably, more than one MoA will be relevant. There are no sound quantitative data to calculate a default ELR_H/ELR_T for substances with intermediate solubility. Therefore, if a default has to be selected, this could only be quantified pragmatically, because of the significant quantitative uncertainties. This does not preclude that adequate interspecies comparisons in elimination are available in individual cases. However, such case-by-case discussions always should consider the uncertainties from translocation of particles to the interstitium in either species.

Finally, for highly soluble particles

i) the quantitative solubility in either epithelial lining fluid, lysosomal fluid, or interstitial fluid has not yet been determined adequately to conclude

definitely on equal elimination rates (i.e. $ELR_H/ELR_T = 1$). Therefore, any chosen solubility value would be rather arbitrary and associated with relevant uncertainty,

ii) again, also soluble particles may be translocated to the interstitium and potential species differences need to be considered for this compartment,

iii) there may be many reasons to deviate from a default with no differences in clearance time, if there are indications of a binding to proteins or other alveolar tissue, which usually are associated with species differences.

It should be emphasised that there may be other clearance mechanisms but just AM mechanical clearance and that species differences are not limited just to AM clearance.

Considering all these bits of information on elimination, it could also be justified to abstain to select a separate ELR_H/ELR_T – ratio. Instead an overall uncertainty factor, which addresses normalisation and elimination simultaneously, without pretending exact scientific background, could be thought of. Such an “aggregate 3 ratios” approach is outlined in Section 7.4.

7 Aggregated HEC-Calculation

7.1 Aggregated HEC-calculation – overview

In the previous Sections we discussed the single four interspecies ratios (weighted breathing volume, normalisation factor, elimination rate, and deposition fraction), which determine HEC. Finally, those ratios are multiplied for aggregation. However, some characteristics are to be acknowledged for this HEC-result:

- Differently from “allometric scaling” for systemic effects, the HEC calculation is significantly influenced by the assumed mode of action. This aspect is briefly discussed in Section 7.2.
- Not always all the four ratios mentioned above can be determined from the experimental data. Therefore, it is sometimes suggested “to take what we have” and neglect those ratios, for which no data are available, i.e., calculate a “partial HEC”. We discuss this proposal in Section 7.3.
- Normalisation factor and elimination rate are interrelated terms, as both are influenced by particle solubility and mode of action. Therefore, an alternative approach is discussed: instead of the traditional “4 ratios approach” for HEC, the human equivalent concentration may also be estimated by an “aggregate 3 ratios approach”, combining the normalisation factor and elimination rate. This alternative is suggested in Section 7.4.
- MPPD 3.04 permits to calculate HEC based on experimental animal inhalation studies with mice. In Section 7.4 we bring together the various data for mice as a starting point to calculate HEC and discuss the consequences.
- As indicated in Section 6.4, elimination rate is different, if translocation to the interstitium is included or excluded from assessment considerations. It is suggested to add an additional assessment factor to address potential translocation to the interstitium. For discussion see Section 7.6.
- The applicability of HEC calculation results may be restricted because of overall uncertainties. Such an applicability constraint is suggested in Section 3.
- Some HEC examples are calculated both, with the “4 ratios approach” and with the “aggregate 3 ratios approach”, to analyse the quantitative consequences of the updates in Sections 3 to 6, and specifically about the application of MPPD 3.04 (Section 7.8).
- Finally, main uncertainties of the HEC approach are summarized in Section 7.9.

7.2 HEC – no isolated precursor step

From the determinants described above (Sections 3 to 6) it is obvious that HEC calculation for respiratory effects in the lower respiratory tract is not a routine procedure, which could be executed adequately without a very good understanding of the mode of action for the observed effects. For HEC calculation qualified information is needed for all four aspects (weighted breathing volume, deposition fraction, normalising factor and retention) for both species. Data compilation and analysis may be more complex than application of allometric caloric demand scaling, which is one step within interspecies extrapolation for systemic effects.

7.3 Partial HEC, if only selected data are available?

As described above (Section 7.2), HEC is a *composed* term. Frequently, we are not able to quantify all of the ratios ((AgV_T / AgV_H), (NF_H/ NF_T), (ELR_H /ELR_T) and (DF_T/ DF_H)) with similar precision or, in cases of high uncertainty, it may even be impossible to quantify some of the ratios at all. Usually there should be sufficiently qualified information on weighted breathing volume ratios and on the deposition faction ratio, but information for normalisation and retention (in combination with solubility and mode of action) may be inadequate.

There are three different options, if such significant uncertainties prevail:

- Option 1: Set HEC/c_T = 1
- Option 2: Use the deposited dose normalised to the alveolar surface and assume no species differences in elimination rates
- Option 3: Combine the ratios, which are uncertain in quantification and apply an aggregate pragmatic assessment factor to cover limited information on normalisation, solubility and retention species differences.

<u>Option 1:</u>
Using HEC/c_T = 1 means that exposure to air concentrations [mg/m³] is assumed to be equipotent for the experimental animals and humans without interspecies corrections. It is suggested to apply this option (HEC/c_T=1), if there is no convincing evidence that one of the ratios with insufficient data needs to be accounted for explicitly. Moreover, after calculating HEC for a set of example substances below (Section 7.8), we found that HEC/c_T = 1 can frequently be selected, if the overall pursued protection goal is moderate. Balancing the various uncertainties in quantification of the single ratios of the HEC formula on the one side and the limited deviations of the complete HEC result from HEC/c_T = 1 on the other side, the assumption that exposure to air concentrations [mg/m³] is equipotent for the experimental animals and humans without corrections for species specific breathing volume, deposition, normalisation and retention may be a reasonable conclusion. If,

however, an elevated protection goal is pursued, an additional assessment factor could be considered.

Option 2:
Some inhalation toxicologists suggest to assume equal elimination rates ($ELR_H/ELR_T = 1$) for particles with some solubility and assume normalisation only to the alveolar surface. This means to use HEC for the deposited dose, instead of the HEC for the retained lung burden. However, this approach is not supported, unless there is substantial evidence that there are no species differences in elimination half-life (which includes not only AM clearance, but also other clearance mechanisms – see Section 6.2) or unless the MoA is clearly linked to the deposited dose instead of the retained dose. We advise against a default selection of $ELR_H/ELR_T = 1$, because in this case the overall probability to create a substantial bias from the "real" HEC is regarded as higher compared to select option 1 (i.e. HEC/c=1).

Option 3:
There is a third option to estimate HEC, if the quality of the data is insufficient to quantify some of the sub-factors and if option 1 (i.e. $HEC/c_T=1$) is regarded not sufficiently protective. This third approach substitutes the single ratios for normalisation and elimination species differences (NF_H/NF_T, ELR_H/ELR_T) by a pragmatic aggregate assessment factor based on the retained dose ("aggregate 3 ratios approach"; Section 7.4).

7.4 Suggested "aggregate 3 ratios approach"

For weighted breathing volumes (Section 3) and deposition fractions (Section 4), default HEC calculations for extrapolation from rat to man can be readily performed with limited additional guidance to be developed. The use of MPPD (version 3.04) is suggested for calculation of deposition fractions.

However, the application of HEC for the ratios (NF_H/NF_T and ELR_H/ELR_T) for interspecies assessments based on rat data (Sections 5 and 6) should be reconsidered, because of significant overall uncertainties. Acknowledging those uncertainties and the complex interrelation of MoA and particle solubility on the one side and the corresponding normalisation and interspecies elimination rate ratio quantification on the other site, we suggest a pragmatic aggregate approach for those ratios in the HEC formula in combination, as outlined below. The problem arising from translocation of particles to the interstitium, which may also affect normalisation and elimination rate is discussed separately (Section 7.6).

Firstly, we recall the quantitative uncertainties on normalisation and elimination rate quantification and the parameters influencing those uncertainties. We combine this information with some suggestions to be reflected in the aggregate approach:

- **Alveolar surface area** is a frequently suggested unit for normalisation. Quantitative figures for NF_H/NF_T ratio are in the range of 140 – 350 (Table

5-1). The difference is partly due to the inclusion or exclusion of the TB region, with no precise generic answer possible, whether the lower TB region should be included or not. Lung surface is regarded as adequate unit for normalisation by many assessors. Only for PSLT particles there are some strong indications that lung surface should not be justified for normalisation. This holds also true for soluble particles. Specifically, for substances readily soluble in alveolar lining fluids the reference to lung surface should be considered. Acknowledging the various uncertainties an average normalisation factor ratio of 250 can be considered based on lung surface. This value is also close to the lung surface estimate by Kuempel et al. (2015).

- Total alveolar macrophage volume is frequently suggested for normalisation with respect to impaired clearance effects by PSLT particles. The range of suggested volume data NFH/NFT ratios is in the range of 280-1110 (Section 5.4; Table 5-2). It should be acknowledged that macrophage clearance and the species differences in particle clearance may also be relevant for non-PSLT particles and soluble particles. Specifically, for substances readily soluble in lysosomal fluids the reference to AM volume should be considered. Acknowledging the various uncertainties an average normalisation factor ratio of 750 can be considered based on total AM volume. This value is also close to the AM volume estimate by Kuempel et al. (2001).

- Therefore, if normalisation is *not clearly only* to the total alveolar macrophage volume, but also the lung surface is to be considered, some factor within the range from 250 to 750 should be use to account for different optimal normalisations depending on the various mode of actions. From this we assign pragmatically
    - a normalisation factor ratio of **600**, if a **major influence** of the alveolar **macrophages** on optimal normalisation is supported, and a
    - normalisation factor ratio of **400**, if a **major influence** of the lung surface on the optimal normalisation is supported.

- The current default in Germany for elimination rate is based on clearance half-life differences for PSLT substances in the rat (40-90d; usually set to 60 days) and in humans (400 d as current default for PSLT). This results in an interspecies default elimination rate ratio of 0.15 (=60/400). However, as has been recently demonstrated, the AM clearance of particles in humans is usually faster (about 300 d). In addition, additional translocation to the interstitium may have to be considered, which would increase the overall elimination half-life in humans well above the former 400 days. As we exclude interstitium translocation in this aggregate 3 ratios HEC approach (focusing on normalisation and elimination) and assign a separate step to this issue (Section 7.6), we can exclude this prolongation in half-life. We, therefore, suggest a slightly increased default interspecies elimination rate of 0.2 (=60/300; Sections 6.4 and 6.9). There are no qualified data on elimination rate species differences based on other MoA apart from PSLT AM clearance impairment. Specifically, for soluble particles, which are not definitely readily eliminated, some species differences are to be expected, but are not yet

sufficiently assessed to provide a sound default elimination rate for interspecies comparisons. For “some” solubility impact on elimination rate is considered ELR_H/ELR_T ratio will be in the range of >0.2 and <1.

From this we apply pragmatically

- ELR_H/ELR_T of 0.25 for substances with low solubility
- ELR_H/ELR_T of 0.5 for substances with intermediate solubility
- ELR_H/ELR_T of 0.8 for substances with high solubility,

to account for species differences in elimination rate, if solubility has “some” impact on elimination duration. We further assume that ELR_H/ELR_T is never equal to the upper or lower limit of the range, i.e. never =0.2 or =1. This means, that minor species differences are assumed also for substances with high solubility and some solubility is also assumed for substances with low solubility.

Aggregate approach (aggregate “normalisation and elimination rate ratio”, NEIR):

We suggest to establish an aggregate term to consider interspecies differences in normalization and elimination and call this ratio **NEIR.** The following situations may be discriminated:

Normalisation *predominantly but not exclusively* to alveolar macrophages volume (i.e., NF-ratio: 600), and intermediate solubility (i.e, ELR –ratio: 0.5): → NEIR will be set to 600 x 0.5 = 300

Normalisation *predominantly but not exclusively* to lung epithelial surface (i.e., NF-ratio: 400), and intermediate solubility (i.e., ELR-ratio: 0.5): → NEIR will be set to 400 x 0.5 = 200

Normalisation *exclusively* to lung epithelial surface (i.e., NF-ratio: 250), and high solubility (i.e, ELR – ratio: 0.8): → NEIR will be set to 250 x 0.8 = 200

Normalisation *exclusively* to alveolar macrophages volume (i.e., NF-ratio: 750), and low solubility (i.e., ELR-ratio: 0.25): → NEIR will be set to 750 x 0.25 ≈ 200

To demonstrate that the last calculation is sufficiently protective, the value for NEIR could also be interpreted as 1110 x 0.2 ≈ 200 for NEIR (PSLT particles), with slightly different selected parameters for AM volume (NF ratio from Table 5-1) or elimination rate (ELR ratio from Section 6.9).
Based on the considerations above, the following Table 7-1 shows case-specific NEIRs, which can be either 200 or 300. However, quantitative figures to classify solubility are to be assigned yet.

Table 7-1 Suggested normalisation and elimination interspecies ratio (NEIR) for normalisation and elimination within HEC calculation (aggregate 3 ratios approach)

NEIR assignment		Lysosomal Solubility	
Solubility in Water	Solubility in ALF	≤# mg/l	>## mg/l
<#mg/l	<#mg/l	200	300
	>#g/l	300	300
>## g/l	irrelevant	Indications of protein binding: yes →300	
		Indications of protein binding: no →200	

Green: dominated by AM-volume normalisation and relevant species differences in clearance
Red: mixed, lung-surface area plus AM-volume normalisation and some species differences in clearance
Yellow: lung surface normalisation (because of high solubility) and no relevant indications of species differences

7.5 Aggregated HEC calculation based on experimental data for mice

From the data reported in Sections 3.4, 4.11, 5.6, and 6.8 it is concluded that no default HEC procedure should be proposed, if experimental data from mouse studies are considered as a starting point for interspecies extrapolations for particles. Usually, experimental data from rat should be preferred, if adverse health effects in the lower respiratory tract from particles are to be assessed. The background of this suggestion are overall uncertainties:

- Considerable variability exists in breathing volume data with current default values on tidal volume and breathing frequency mainly just based on two strains (BALB/c- and B6C3F$_1$-mice).
- No direct functional residual capacity data (FRC) for mice were available, but were indirectly derived from rat data (Asgharian et al. 2014; Hsieh et al. 1999).
- There is a dramatic difference, if either normalised to the lung epithelial surface or to the alveolar macrophage volume, with no unambiguous decision criteria, which of the normalisation references should be applied (Section 5.6).
- Documented alveolar surface area data for mice are highly divergent (Section 5.6).
- Few data on representative elimination half-life, specifically, if particle clearance is not only by AM-mucociliary clearance.
- No specific data are available for inhomogeneity in deposition and retention for mice (“hot spots”).

- The exclusion or inclusion of the TB-region in addition to the PU-region may be even more relevant for mice compared to rats, because of the generally high particle fraction, deposited in the TB-region in this species.

Uncertainties are also shown in quantitative example calculations (Section 7.8.4).

7.6 Translocation to the interstitium: suggested separate sub-factor in HEC default calculations

As indicated above (Section 6.4), species differences due to translocation of particles to the lung interstitium are not covered in most former HEC approaches and in the suggested default calculation procedure in this report, so far. However, a separate factor to address this issue is suggested in this Section.
From the work by Kuempel et al. (2015; 2001) and by Gregoratto et al. (2010, 2011) it is concluded that interspecies differences from clearance of particle to the lung interstitium should be included in HEC calculations, if relevant. However, there are major uncertainties, i) whether interstitial effects in humans are sufficiently represented in animal studies, i.e. whether the rat is a suitable model animal to provide information on interstitial effects in humans, ii) whether the interstitial effects observed in rats at relatively low exposures only with soluble metal particles are representing relevant effects in human exposure and iii) whether interstitial effects at "overload doses" in rats of PSLT substances are quantitatively comparable to interstitial effects in humans (Bos et al. 2019).
However, there is clear evidence that elimination half-life of particles in humans is significantly increased, if the interstitium is included as part of the pulmonary region (multi-phase elimination). Further discussion on interstitial effects from particles, translocation and species differences are provided in Section 6.4. In order to cover respective differences with respect to deposition, normalisation, and clearance, we suggest to apply an additional

interspecies sub-factor of 0.5 (IIF = interstitium interspecies factor)

within the HEC-procedure, if there is any specific indication, that interstitium effects may be relevant in either species for a certain particle assessment.

"Any specific indication" means, that either in humans (qualitative or quantitative) fibrotic effects are associated with exposure to this particle or if fibrotic effects or (malignant or benign) tumours in interstitial tissue are observed in laboratory animals above control background rate. It is not necessary that such interstitial effects are shown in both species in interspecies comparisons, to justify the need of an IIF.

The suggested factor of 0.5 is taken from an analysis by Kuempel et al. (2015). The authors compared different models and retained lung doses, which differed, whether the interstitium was explicitly addressed or excluded, and found "the estimates from these models differed by a factor of 2-3 with the interstitial-sequestration model predicting lower air borne concentrations associated with the working lifetime

retained lung burden" (Section 6.4). Because the inverse of this factor (2-3) is applied within the HEC-formula, IIF is set to ½ = 0.5.

Therefore, the default HEC-formula within the **4 ratios approach** or the **aggregate 3 ratios approach** would be modified according the suggestions from Section 7.4

$HEC/c_T = (AgV_T / AgV_H) \times (NF_H / NF_T) \times (ELR_H / ELR_T) \times (DF_T / DF_H) \times IIF$, and

$HEC/c_T = (AgV_T / AgV_H) \times NEIR \times IIF \times (DF_T / DF_H)$

with IIF only, if applicable. Note that the **4 ratios approach** therefore **changes to 5 multipliers** and the **aggregate 3 ratios approach changes to 4 multipliers**, without changes in the terminology.

7.7 Range constraints for the HEC-approach

Most of the aspects to be covered by the HEC-calculation have been addressed above (Sections 7.2 to 7.6). However, some major uncertainties remain:

- the consequences of inhomogeneous distribution (hot spots),
- the possible hygroscopic particle growth effects,
- the potential difference in particle size distribution in human vs. experimental animal exposure.

There are no qualified approaches available, how to address the mentioned additional uncertainties quantitatively. We therefore suggest to limit any default HEC – applicability to an upper HEC/c= 1, because of overall considerations on protective assessment factors.

We therefore suggest to calculate HEC within the **4 ratios approach**:

a) within the 4 ratios approach:

HEC/c_T

$= (AgV_T / AgV_H) \times (NF_H / NF_T) \times (ELR_H / ELR_T) \times (DF_T / DF_H) \times IIF$, if $HEC/c_T < 1$ and

$HEC/c_T = 1$,

if $(AgV_T / AgV_H) \times (NF_H / NF_T) \times (ELR_H / ELR_T) \times (DF_T / DF_H) \times IIF$, if $HEC/c_T \geq 1$

b) or within the **aggregate 3 ratios approach**:

$HEC/c_T = (AgV_T / AgV_H) \times NEIR \times IIF \times (DF_T / DF_H)$, if $HEC/c_T < 1$ and

$HEC/c_T = 1$, if $(AgV_T / AgV_H) \times NEIR \times IIF \times (DF_T / DF_H)$, if $HEC/c_T \geq 1$

Note that this is a suggestion for default HEC calculations. It is always accepted to deviate from default in case of qualified data and case-by-case justification.

7.8 Some examples for HEC

7.8.1 HEC - calculation from rat data for PSLT substances: titanium dioxide

This example is selected in order to compare the results from HEC calculation by Pauluhn (2011b) and by Hartwig (2012) with the respective updated considerations, as documented and discussed in the present report. For titanium dioxide only the study by Muhle et al. (1991) was used for parameter specification. It is not intended to present an overall assessment on titanium dioxide, but only to compare the HEC-transformed NOAEC in rats with the corresponding human NAEC.

Input parameters:

NOAEC (rat) = 5 mg/m³

MMAD: 1.1 µm

GSD: 1.6

Density: 4.3

Exposure 6h/d; 5d/w; 2 years

Strain: F344 rats

Body weight: no data provided; used: 370 g (F344 male, 12 months, (Mauderly 1986))

MPPD calculation of deposition: version 3.04

Results:

1.) AgV_T/AgV_H

Non-default:

Breathing scenario rat in the respective original study: whole body exposure (**not:** nose only)

Breathing frequency: 109.0787 bpm according to MPPD exposure for given body weight

Tidal volume: 2.59 mL according to MPPD exposure for given body weight

⇨ 101.706 l/d Breathing volume = 0.102 m³/d

- ⇨ 0.102 m³/d x 5/7 = 0.073 (chronic weighted breathing volume, rat)
- ⇨ 0.073 (this calculation, rat)/ 6.57 (average human) = 0.011 = AgV_T/AgV_H

The calculated AgV_T/AgV_H differs only slightly (factor ≈ 1.4) from the former value (0.008).

2.) DF_T/DF_H

DF_T/DF_H calculated as shown in Table 7-2. We limit subsequent calculations to the pulmonary region, which results in a DF_T/DF_H-ratio of 0.692. This ratio is close to identical to the parallel calculation with MPPD 2.11 (MPPD 2.11; PU-DF_T/DF_H = 0.695; data not shown). Note, however, that the two calculations would lead to some difference, if the TB-region would be included, i.e. MPPD 3.04: PU+TB-DF_T/DF_H = 0.89 vs. MPPD 2.11: PU+TB-DF_T/DF_H = 0.68). There probably is some influence in the deposition calculation with “whole body exposure” vs. “nose only” exposure; this distinction is only possible in MPPD 3.04 and is based on a different allometric calculation compared to the “nose only” scenario. MPPD 2.11 does not permit to differentiate “nose only” vs. “whole body” – exposure.

Table 7-2 Example calculation of deposition fractions and DF_T/DF_H for Titanium dioxide (data from Muhle et al., (1991)), MPPD 3.04

	MPPD 3.04		
	PU	TB	PU+TB
DF_H	0.1046	0.0404	0.1450
DF_T	0.0724	0.0573	0.1297
DF_T/DF_H	0.6922	1.14183	0.8945

3.) Normalisation

In most former approaches, PSLT particles were normalised to the alveolar macrophage volume. This normalisation reference is confirmed for PSLT in this report. According to Pauluhn (2011b) (Table 5-2, this report), a normalising factor NFH/NFT of ≈ 1110 would be used. Instead, some other data provided in Table 5-2 within the range of 278 and 1110. For the purpose of this example calculation we use an average value of 700 ((278+1110)/2≈700) for the 4 ratios approach in HEC calculation.

4.) Retention

The traditional retention factor ratio is derived from the elimination half-life from the lung in humans (400 d) and the respective value for rats (60 d). However, as discussed in Sections 6.4 and 6.9, human lung elimination half-life is reduced, if the interstitium compartment is excluded as part of the target organ (lung), with ELR_H/ELR_T = 0.2 instead of ELR_H/ELR_T = 0.15 for PSLT-particles. This is not the

case for titanium dioxide, where interstitium effects were observed in the rat (although only at higher exposures). Therefore, according to Section 7.6, an additional factor of 2 (IIF= 0.5) is suggested to be used supplementary in both, the “4 ratios approach” and the “aggregate 3 ratios approach”.

5.) Overall HEC

This example assessment provides a HEC for titanium dioxide, based on the study by Muhle et al. (1991). If each single factor is quantified (**4 ratios approach**) this results in the following HEC:

HEC= 0.011 x 0.692 x 700 x 0.2 x 0.5 x c_T = **0.53** x c_T
If the **aggregate 3 ratios approach** (Section 7.4) with a NEIR of 200 is used (as depicted from Table 7-1),

HEC= 0.011 x 0.692 x 200 x 0.5 x c_T = **0.76** x c_T

With a $NOAEC_{rat}$ of 5 mg/m³ in experimental animals these two calculations result in similar values of $NAEC_{HEC}$= 2.66 (4 ratios approach) – 3.81 mg/m³ (3 ratios approach).

6.) Discussion

Calculations for titanium dioxide based on the identical study performed by Hartwig (2012) resulted in a $NAEC_{HEC}$= 1.06 mg/m³ (rounded to 1.1 mg/m³; Section 5.2). Hartwig used the lung surface for normalisation, which resulted in smaller values. Minor additional differences are due to the additional inclusion of IIF because of the interstitium effects. Detailed MPPD reports were not available and an earlier version of MPPD was used by Hartwig (2012). These values (2.66-3.81 mg/m³) are slightly higher than the threshold calculated with approach B in Section 5.2 (2.15 mg/m³ for a density of 4.3), but in good agreement with the former calculation. It also considers possible interstitium effects, which were not included previously. The **aggregate 3 ratios approach** provides similar results as the more detailed **4 ratios approach**.

7.8.2 HEC – calculation from rat data for water-soluble particles: cobalt sulfate

This example illustrates the use of MPPD version 3.04 in combination with further updates in HEC calculation discussed in the present report for a substance, which is soluble in water (337.4 g/Liter at 20°C), and also soluble in lysosomal fluid and in epithelial lining fluid (AGS 2017). The experimental data, as reported by NTP (1991, 1998) for cobalt sulfate heptahydrate were used for parameter specification and transformed to cobalt sulfate, if applicable. It is not intended to present an overall assessment on cobalt sulfate, but only to compare the HEC-transformed NOAEC in rats with the corresponding human NAEC.

Input parameters:

NOAEC (rat) = 67 µg/m³ (no pulmonary adverse effects in a subchronic study) (NTP 1991)

MMAD: 1.4 µm (NTP 1991)

GSD: 2.1 (NTP 1991)

Density: 2.0 g/cm³ (Zalkin et al. 1961)

Exposure 6h/d; 5d/w

Strain: F344 rats

Body weight: 434 grams (NTP 1998)

MPPD calculation of deposition: version 3.04

Results:

1.) AgV_T/AgV_H

Breathing frequency: 166 (usually for SD-rats, adopted)

Tidal volume: 3.24 according to MPPD, exposure for given body weight

- ⇨ 193622 l/d Breathing volume = 0.193 m³/d
- ⇨ 0.193/6.57 = 0.029
- ⇨ 0.029 x 5/7 = 0.021 = AgV_T/AgV_H

The calculated AgV_T/AgV_H differs by a factor of 2.6 from the former value (0.008)

2.) DF_T/DF_H

DF_T/DF_H was calculated as shown in Table 7-3. We limit subsequent calculations to the pulmonary region, which results in a DF_T/DF_H-ratio of 0.33. Uncertainties in deposition come from hygroscopic growth for water- soluble particles, but were not considered in this calculation.

Table 7-3 Example calculation of deposition fractions and DF_T/DF_H for cobalt sulfate (data from NTP (1991, 1998)), MPPD 3.04

	MPPD 3.04		
	PU	TB	PU+TB
DF_H	0.0944	0.0354	0.1298
DF_T	0.0313	0.0313	0.0626
DF_T/DF_H	0.33	0.88	0.48

3.) Normalisation

In most former approaches, normalisation ratio for water-soluble particles would have been to the alveolar surface area (NF_H/NF_T). Quantitative figures (Table 5-1) are in the range from 140 to 349 for NF_H/NF_T. For the purpose of this example, we use an average value of 250 ((349+140)/2 ≈ 250) for the **4 ratios approach** HEC-calculation.

4.) Retention

The retention factor in the updated HEC-approach is derived from the elimination half-life from the lung in humans (300d) and the respective value for rats (60d). However, for water- soluble substances, no species differences are assumed ($ELR_H/ELR_T=1$). However, as discussed in Sections 6.4 and 6.9, species effects in retention are expected, if the interstitium compartment is included. This is the case for cobalt sulfate, where interstitium effects were observed in the rat (NTP 1998). Therefore, according to Section 7.6, an additional factor of 2 (IIF= 0.5) is suggested for both, the **4 ratios approach** and the **aggregate 3 ratios approach**.

However, although water-solubility of cobalt sulfate is high, there is evidence that the substance binds to proteins: "... *in vivo*, the bioavailability of free Co(II) is expected to be relatively limited, because these cations precipitate in the presence of physiological concentrations of phosphates" (Paustenbach et al. 2013) and also supported by further *in vitro* –observations (Stopford et al. 2003). This is not considered in the **4 ratios approach**. If the **aggregate 3 ratios approach** is used, NEIR of 300 is suggested for such substances according to the scheme in Section 7.4.

5.) Overall HEC

This example assessment provides a HEC for cobalt sulfate, based on the study by NTP (1991, 1998). If each single factor is quantified (**4 ratios approach**) this results in the following HEC:

HEC= 0.021 x 0.33 x 250 x 1 x 0.5 x c_T = **0.86** x c_T

If the **aggregate 3 ratios approach** of Section 7.4 with NEIR = 300 is used,

HEC= 0.021 x 0.33 x 300 x 0.5 x c_T = **1.04** x c_T, changed to **$HEC/c_T = 1$** (according to Section 7.7)

According to Section 7.7, $HEC/c_T > 1$ are not permitted in the suggested update, because of overall uncertainties. Therefore, in the aggregate approach, HEC/c_T is set to 1 (and $NOAEC_{rat} = NAEC_{HEC}$).

In consequence, with the 4 ratios approach

HEC = 0.86 x 67 µg/m³ = 58 µg/m³,

With the 3 ratios approach, the NOAEC rat is maintained as the NOAEC for humans:

HEC = 67 µg/m³.

6.) Discussion

Note that the **4 ratios approach** and the **aggregate 3 ratios approach** come up with very similar results close to HEC/c_T=1.

7.8.3 HEC – calculation from rat data for particles with lysosomal solubility: cobalt metal

This example illustrates the use of MPPD version 3.04 in combination with further updates in HEC calculation discussed in the present report for a substance, which is poorly soluble in water (2.9 mg/Liter at 20°C) and in epithelial lining fluid (4.8 % solubility for extra fine particles at pH 7.4), but soluble in (macrophage) lysosomal fluid (92.4 % solubility for extra fine particles, at pH 4.5) (AGS 2017). This example is performed with parameters for rats and can be compared to the example below (Section 7.8.4) with mice. The experimental data, as reported by NTP (2014) for cobalt metal were used for parameter specification. It is not intended to present an overall assessment on cobalt metal, but only to compare the HEC-transformed NOAEC in rats with the corresponding human NAEC.

Input parameters:

LOAEC (rat) = 1.25 mg/m³

MMAD: 1.8 µm

GSD: 1.7

Density: 8.81 g/cm³

Exposure 6h/d; 5d/w

Strain: F344 rats

Body weight: 434 grams (assumed identical to the cobalt sulfate study, within the range of body weights, male rats, for cobalt metal)

MPPD –calculation of deposition: version 3.04

Results:

1.) AgV_T/AgV_H

Breathing frequency: 166 (usually for SD-rats, adopted)

Tidal volume: 3.24 mL according to MPPD, exposure for given body weight

- ⇨ 193622 mL/d Breathing volume = 0.193 m³/d
- ⇨ 0.193/6.57 = 0.029
- ⇨ 0.029 x 5/7 = 0.021 = AgV_T/AgV_H

The calculated AgV_T/AgV_H differs by a factor of 2.6 from the former value (0.008).

2.) DF_T/DF_H

DFT/DFH was calculated as shown in Table 7-4. We limit subsequent calculations to the pulmonary region, which results in a DFT/DFH factor of 0.2. Uncertainties in deposition come from hygroscopic growth for water-soluble particles but were not considered in this calculation.

Table 7-4 Example calculation of deposition fractions and DF_T/DF_H for cobalt metal (data from NTP (2014)), MPPD 3.04

	MPPD 3.04		
	PU	TB	PU+TB
DF_H	0.1147	0.0372	0.1519
DF_T	0.0227	0.0346	0.0573
DF_T/DF_H	0.20	0.93	0.38

3.) Normalisation

Because of the low water solubility of cobalt metal particles, in most former approaches, normalisation would have been to the total alveolar macrophage volume (NFH/NFT = 670 according to Kuempel et al. (2015) or NFH/NFT = 1110, according to Pauluhn (2011b) (Table 5-2). However, due to the significant lysosomal solubility (e.g., in the macrophages), the macrophage volume may not be decisive for normalisation. Hence, also normalisation to the lung surface should be considered (NFH/NFT = e.g., 150; Table 5-1). Within the 4 ratios approach we selected NFH/NFT= 600 ((150+1110)/2≈600) for example calculations below.

4.) Retention

The updated retention factor is derived from the elimination half-life from the lung in humans (300d) and the respective value for rats (60d), resulting in ELR_H/ERL_T=0.20. However, for substances with medium solubility in the German approach, a doubling of ELR_H/ERL_T is suggested. We take account of the lysosomal solubility by reducing the interspecies difference in elimination, i.e. by doubling from ELR_H/ERL_T from 0.20 to 0.4. However, as discussed in Sections 6.4 and 6.9, species effects in retention are expected, if the interstitium compartment is included. This is the case for cobalt metal, where interstitium effects were observed in the rat, although only at higher doses. Therefore, according to Section 7.6, an additional factor of 2 (IIF= 0.5) is suggested for both, the **4 ratios approach** and the **aggregate 3 ratios approach**.

In the **aggregate 3 ratios approach** factors for normalisation and for retention are combined to a single NEIR of 300 (Section 7.4).

5.) Overall HEC

This example assessment provides a HEC for cobalt metal, based on the study by NTP (2014). If each single factor is quantified (**4 ratios approach**) this results in the following HEC:

HEC= 0.021 x 0.2 x 600 x 0.4 x 0.5 x c_T = **0.5** x c_T

If the **aggregate 3 ratios approach** of Section 7.4 is used,

HEC= 0.021 x 0.2 x 300 x 0.5 x c_T = **0.63** x c_T

With a $LOAEC_{rat}$ of 1.25 mg/m³ these two calculations result in similar values of $LOAEC_{HEC}$= 0.63 (4 ratios approach) – 0.79 mg/m³ (3 ratios approach).

6.) Discussion

Again this example calculation provided $LAEL_{HEC}$, which were similar to the $LOAEC_{rat}$ in the animal study, because of a HEC/c_T = 0.5 (**4 ratios approach**) and a very similar HEC/c_T of 0.63 (**aggregate 3 ratios approach**). A small deposition fraction of only ≈ 2 % in the rat resulted in large interspecies difference in deposition (DF_T/DF_H = 0.2). However, the factors used for normalisation and retention in the 4 ratios approach have been selected only as example quantification for the purposes of this calculation and are not agreed by expert evaluations.

7.8.4 HEC – calculation from mice data for particles with lysosomal solubility: cobalt metal

This example illustrates the use of MPPD version 3.04 in combination with further updates in HEC calculation discussed in the present report for a substance, which is poorly soluble in water (2.9 mg/Liter at 20°C) and in epithelial lining fluid (4.8 % solubility for extra fine particles at pH 7.4), but soluble in lysosomal fluid (92.4 % solubility for extra fine particles, at pH 4.5) (AGS 2017). This example is performed with parameters for mice and can be compared to the example above (Section 7.8.3) with rats. The experimental data, as reported by NTP (2014) for cobalt metal were used for parameter specification. It is not intended to present an overall assessment on cobalt metal, but only to compare the HEC-transformed NOAEC in rats with the corresponding human NAEC.

Input parameters:

LOAEC (mice) = 1.25 mg/m³

MMAD: 1.8 µm

GSD: 1.7

Density: 8.81 g/cm³

Exposure 6h/d; 5d/w

Strain: B6C3F1-mice

Body weight: 47.6 grams (average, male, after 52 weeks, NTP (2014))

MPPD –calculation of deposition: version 3.04

Results:

1.) AgV_T/AgV_H

Breathing frequency: 243 (from MPPD-exposure for given body weight, B6C3F1-mice)

Tidal volume: 0.262 mL (Note that the tidal volume, as suggested by MPPD template, has been corrected according to the allometric formula provided in Section 3.4)

- ⇨ 22920 mL/d Breathing volume = 0.0229 m^3/d
- ⇨ 0.0229/6.57 = 0.0035
- ⇨ 0.0035 x 5/7 = 0.0025 = AgV_T/AgV_H

The calculated AgV_T/AgV_H differs only marginally (0.0025 vs. 0.0037) from the one derived in Section 3.4.

2.) DF_T/DF_H

DF_T/DF_H was calculated as shown in Table 7-5. We limit subsequent calculations to the pulmonary region, which results in a DF_T/DF_H-factor of 0.39. There is some influence in the deposition calculation by the assumed breathing pattern (breathing frequency, tidal volume) calculated from allometric formula by MPPD based on the given body weight. Note, that the tidal volume, as suggested by MPPD template, has been corrected according to the allometric formula provided in Section 3.4.

Table 7-5 Example calculation of deposition fractions in the mouse lung and DF_T/DF_H for cobalt metal (data from NTP (2014)), MPPD 3.04

	MPPD 3.04		
	PU	TB	PU+TB
DF_H	0.1147	0.0372	0.1519
DF_T	0.0448	0.0219	0.0667
DF_T/DF_H	0.39	0.59	0.44

3.) Normalisation

Because of the low water solubility of cobalt metal particles, in most former approaches, normalisation would have been to the total alveolar macrophage volume. However, because of an assumable high solubility in lysosomal fluid, macrophage loading is assumed to be limited and effects may arise from macrophage damage as well as from damages of epithelia of the lung cells. For rats,

we calculated NF_H/NF_T, influenced by both, macrophage volume and lunge surface. Such a factor is not available for mice and should not be proposed without more qualified data. Therefore we provide NF_H/NF_T for both options (option 1: total macrophage volume; option 2: lung epithelial surface), to show the range of results for comparison.

4.) Retention

The updated retention factor is derived from the elimination half-life from the lung in humans (300d) and the respective value for mice (60d), resulting in ELR_H/ERL_T=0.2. However, for substances with medium solubility in the German approach, a doubling of ELR_H/ERL_T is suggested. We take account of the lysosomal solubility by reducing the interspecies difference in elimination, i.e. by doubling from ELR_H/ERL_T from 0.2 to 0.4 (both options). However, as discussed in Sections 6.4 and 6.9, species effects in retention are expected, if the interstitium compartment is included. This is the case for cobalt metal, where interstitium effects were observed in the mouse.

5.) Overall HEC

This example assessment provides a HEC for cobalt metal, based on the study by NTP (2014). This results in the following HEC (option 1: normalisation to the alveolar macrophage volume of ≈ 13000 according to Stone et al. (1992); Section 5.6):

HEC= 0.0025 x 0.39 x 13000 x 0.4 x 0.5 x c_T = **2.54** x c_T,
changed to **HEC/c_T= 1** (according to Section 7.7)

If option 2 (normalisation to the lung epithelial surface based on data from Hsieh et al. (1999) were regarded more appropriate, HEC will be calculated accordingly:

HEC= 0.0025 x 0.39 x 723 x 0.4 x 0.5 x c_T = **0.14** x c_T

$LOAEC_{HEC}$

= 1.25 mg/m³ x 1 =1.25 mg/m³ (option 1) vs. 1.25 x 0.14 ≈ 0.2 mg/m³ (option 2)

For mice, the **aggregate 3 ratios approach** has not been developed due to overall uncertainties (Section 7.5).

6.) Discussion

This last example demonstrates that HEC calculations with mice ($LOAEC_{HEC}$ = 0.2-1.25 mg/m³) do not necessarily contradict HEC from rat ($LOAEC_{HEC}$ = 0.63-0.79 mg/m³) Section 7.8.3). However, a significantly larger range of potential HEC values is derived from mice data due to uncertainties in normalisation. This supports our conclusion that HEC calculations from mice data are not sufficiently validated to be useful in regulatory standard setting.

7.8.5 Conclusions from the examples

The examples calculated above demonstrate that HEC/c_T is always greater than 0.5, if rat data were the starting point. Even though there may be examples with smaller HEC values, such cases are rarely expected. For particles with low solubility and/or relevant contributions of the alveolar macrophages clearance to the MoA the values become even closer to 1. If, however, other equally defendable values were chosen to quantify normalisation and elimination species differences, the influence of HEC on the final result could be substantial.

Differences between the 4 ratios approach and the aggregate 3 ratios approach are rather small: this supports to apply the aggregate 3 ratios approach, as this calculation does not pretend to discriminate the various influences on HEC as precisely as it may be erroneously concluded from the 4 ratios approach.

Considering the limited influence and the high uncertainty of HEC and the elaborate calculations to perform such HEC assessments, it could also be decided to abstain from any HEC calculation, i.e., select HEC/c= 1, if only a moderate protection goal is regarded to be sufficient. If, however, a higher protection goal is regarded to be adequate, even a simple “assessment factor” could serve to fulfil this requirement and the time-consuming aggregate 3 ratios approach could also be waived. This does not preclude the use of all the HEC elements for interspecies extrapolations in case of non-default assessments. If, for example, there are indications of relevant species differences because of contradicting human and animal data, this situation could be analysed using the complete steps of HEC.

The examples also demonstrated the considerably larger range of HEC results for mice data compared to the rat. This is in agreement with our general reluctance to support HEC calculations based on mice data, as, currently, the uncertainties may be too high and the number of qualified data for mice may not be sufficient.

7.9 Summary of uncertainties in the HEC approach

When deriving an occupational exposure limit (OEL) for particles, interspecies extrapolation from rodents to men is frequently necessary. Therefore, it is important to define human exposure levels, which are regarded equivalent to the exposure level of the starting point (the experimental animal inhalation study). Thus, HEC calculation is a major element of interspecies extrapolation for particle exposure. However, as described more closely in Sections 3 to 6, there are several relevant uncertainties to be acknowledged, which are summarized below:

- Quantification of **weighted breathing volumes** should consider strain-specific data, because body weight significantly influences the breathing volume. Uncertainties come from discrepant allometric scaling procedures (MPPD vs. OEHHA), which do not equally cover data from some relevant rat strains and from high variability in breathing volume.

- For **deposition**, some relevant uncertainties arise from potential particle solubility (particle growth for hygroscopic particles), from the reference deposition area (pulmonary only vs. pulmonary plus tracheobronchial sites), and from the inhomogeneity of deposition (hot spots). The different versions of MPPD software may lead to significant discrepancies in the calculated deposition fraction ratio. However, this uncertainty is mainly relevant at larger particle sizes, beyond the default applicability range of the HEC approach.
- There are major uncertainties from the interrelationship between particle solubility and mode of action (MoA) on the one side and the unit for **dose metrics and normalisation** on the other side. It is not always evident, whether the primary site of deposition is also the critical target, relevant for normalisation, or if the site of secondary reactions is the more adequate reference point for normalisation and dose metrics. Frequently, more than one single MoA may be relevant, each with different optimal normalisation unit and a different optimal dose metric. Even if the critical target cells (e.g. lung alveolar epithelium cells or alveolar macrophages) can be identified, there are some discrepancies to quantify the normalisation factor.
- Similar uncertainties are obvious for interspecies clearance and **elimination rate** differences, when the retained dose is the starting point. Earlier procedures only considered alveolar macrophage mucociliary clearance as relevant for interspecies differences. However, translocation to the interstitium needs also to be considered as one of the significant elimination routes with potential interspecies differences. Similarly, earlier approaches assumed that there would be no significant species difference for highly soluble particles in elimination. However, different solubility in the various lung fluids may influence MoA, retention time, and migration to intracellular regions or extracellular effects. If, for example alveolar macrophages are involved in lysis or transport of particles, species differences can be expected. Highly soluble particles may be bound to proteins with consequences in MoA and clearance mechanism with subsequent species differences. Again, inhomogeneity in elimination time due to local hot spot accumulation is not adequately covered by the HEC-calculation procedure.
- Further uncertainties are linked to the default applicability range for default HEC calculations: the experimental animal study should be performed with **particles sizes** in the range of [0.5-2] µm MMAD. Therefore, there may be many assessments and data, where this default needs to be modified with no detailed generic guidance. Specifically, specific considerations apart from the standard HEC-calculation may be necessary for nano-sized particles or agglomerates. It needs further elaborations or, at least transparent discussion of additional recent data, to ensure that OELs calculated for the range of micro-sized particles are equally applicable to the nano-sized agglomerates below 0.5 µm MMAD-equivalent. Moreover, HEC-linked standard setting for particles refer to *all respirable* particle sizes, where significant differences can be observed between the experimental particle size distribution and the size distribution relevant in the human workplace exposure scenario. As particle size dependent deposition fraction and interspecies deposition fraction ratio are not proportional to the size specific dose response relationship for adverse

effects in the animal study, and as animals are usually only studied at a single particle size distribution, this difference in exposure sizes contributes to overall uncertainty. Interspecies extrapolation from small particle sizes in the animal study is not always protective, if humans are exposed to larger respirable particles.

In conclusion, calculation of HEC is one element within the framework of standard setting for particle effects, covering potential toxicokinetic differences between species. However, as shown with the list above, there remain substantial uncertainties with the application of the HEC procedure, which need to be addressed when deriving OELs for particulate substances. Further, other elements of interspecies extrapolation such as potential differences in toxicodynamics, which are outside the scope of this report, should also be addressed.

References

AGS, Ausschuss für Gefahrstoffe. (2013). Leitfaden zur Quantifizierung stoffspezifischer Expositions-Risiko-Beziehungen und von Risikokonzentrationen bei Exposition gegenüber krebserzeugenden Gefahrstoffen am Arbeitsplatz, (Anlage 3 zu TRGS 910). Version N10, Stand: 15.09.2013, Ausarbeitung durch den Arbeitskreis Risikoableitung im UA III des AGS. http://www.baua.de/de/Themen-von-A-Z/Gefahrstoffe/TRGS/TRGS-910.html.

AGS, Ausschuss für Gefahrstoffe. (2014). Technische Regeln für Gefahrstoffe – Arbeitsplatzgrenzwerte (TRGS 900). Ausgabe: Januar 2006. BArBl Heft 1/2006 S. 41-55. Zuletzt geändert und ergänzt: GMBl 2014 S. 1312-1313 v. 28.10.2014 [Nr. 64], Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. http://www.baua.de/cae/servlet/contentblob/666762/publicationFile/55588/TRGS-900.pdf.

AGS, Ausschuss für Gefahrstoffe. (2015). Begründung zum Beurteilungsmaßstab NanoGBS. Ausgabe: Juni 2015. Stand: Mai 2015, https://www.baua.de/DE/Aufgaben/Geschaeftsfuehrung-von-Ausschuessen/AGS/pdf/A-Staub.pdf?__blob=publicationFile&v=2.

AGS, Ausschuss für Gefahrstoffe. (2017). Begründung zu Cobalt und Verbindungen in TRGS 910 (Fassung v. 13.10.2017). Ausgabe: Juni 2017. Stand: November 2012. Cobalt-Metall und anorganische Cobalt-Verbindungen (gilt nur für A-Staub), Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. https://www.baua.de/DE/Aufgaben/Geschaeftsfuehrung-von-Ausschuessen/AGS/AGS_node.html.

Ali, M., Gutting, B.W., van Hoek, M.L. (2017). Multiple path particle dosimetry for prediction of mouse lung deposition of nanoaerosol particles. International Journal of Advanced Research in Engineering and Technology (IJARET) 8, 10-20.

Anjilvel, S., Asgharian, B. (1995). A multiple-path model of particle deposition in the rat lung. Fundamental and Applied Toxicology 28, 41-50.

ARA, Applied Research Associates, Inc. (2018). Multiple-Path Particle Dosimetry Model (MPPD v 3.04), https://www.ara.com/mppd/.

Asgharian, B. (2018). Multiple Path Particle Dosimetry (MPPD) model and its Applications, Applied Research Associates, Inc., Raleigh, NC. https://www.piscltd.org.uk/wp-content/uploads/2018/08/2018.08.01_Bahman-Asgarian_PETA-ISC-webinar-1.pdf.

Asgharian, B. (2019). Multiple Path Particle Dosimetry (MPPD) model and its Applications, Applied Research Associates, Inc., Raleigh, NC. https://www.piscltd.org.uk/wp-content/uploads/2019/02/2019.02.14_ARA_Bahman-Asgharian.pdf.

Asgharian, B., Anjilvel, S. (1998). A multiple-path model of fiber deposition in the rat lung. Toxicological Sciences 44, 80-86.

Asgharian, B., Hofmann, W., Bergmann, R. (2001). Particle deposition in a multiple-path model of the human lung. Aerosol Science and Technology 34, 332-339.

Asgharian, B., Price, O.T., Oldham, M., Chen, L.C., Saunders, E.L., Gordon, T., Mikheev, V.B., Minard, K.R., Teeguarden, J.G. (2014). Computational modeling of nanoscale and microscale particle deposition, retention and dosimetry in the mouse respiratory tract. Inhalation Toxicology 26, 829-842.

Asgharian, B., Price, O.T., Schroeter, J.D., Kimbell, J.S., Singal, M. (2012). A lung dosimetry model of vapor uptake and tissue disposition. Inhalation Toxicology 24, 182-193.

Bailey, M.R., Ansoborlo, E., Guilmette, R.A., Paquet, F. (2007). Updating the ICRP human respiratory tract model. Radiation Protection Dosimetry 127, 31-34.

Bailey, M.R., Kreyling, W.G., Andre, S., Batchelor, A., Collier, C.G., Drosselmeyer, E., Ferron, G.A., Foster, P., Haider, B., Hodgson, A., Masse, R., Metivier, H., Morgan, A., Müller, H.L., Patrichk, G., Pearman, I., Pickering, S., Ramsden, D., Stirling, C., Talbot, R.J. (1989). An interspecies comparison of the lung clearance of inhaled monodisperse cobalt oxide particles - Part I: Objectives and summary of results. Journal of Aerosol Science 20, 169-188.

Balásházy, I., Hofmann, W., Heistracher, T. (2003). Local particle deposition patterns may play a key role in the development of lung cancer. Journal of Applied Physiology 94, 1719-1725.

Benson, J.M., Chang, I.-Y., Cheng, Y.S., Hahn, F.F., Kennedy, C.H., Barr, E.B., Maples, K.R., Snipes, M.B. (1995). Particle clearance and histopathology in lungs of F344/N rats and B6C3F$_1$ mice inhaling nickel oxide or nickel sulfate. Fundamental and Applied Toxicology 28, 232-244.

Bevan, R.J., Kreiling, R., Levy, L.S., Warheit, D.B. (2018). Toxicity testing of poorly soluble particles, lung overload and lung cancer. Regulatory Toxicology and Pharmacology 100, 80-91.

Beyersmann, D., Hartwig, A. (2008). Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms. Archives of Toxicology 82, 493-512.

Bos, P.M.J., Gosens, I., Geraets, L., Delmaar, C., Cassee, F.R. (2019). Pulmonary toxicity in rats following inhalation exposure to poorly soluble particles: The issue of impaired clearance and the relevance for human health hazard and risk assessment. Regulatory Toxicology and Pharmacology 109, Article 104498.

Braakhuis, H.M., Park, M.V., Gosens, I., De Jong, W.H., Cassee, F.R. (2014). Physicochemical characteristics of nanomaterials that affect pulmonary inflammation. Particle and Fibre Toxicology 11:18, 25pp.

Brown, J.S., Wilson, W.E., Grant, L.D. (2005). Dosimetric comparisons of particle deposition and retention in rats and humans. Inhalation Toxicology 17, 355-385.

Bruch, J. (2013). Die pulmonale Oberfläche – Vergleich Ratte und Mensch –. AGSpublik 8 April 2013, Dortmund, Universitätsklinikum Essen. https://www.baua.de/DE/Aufgaben/Geschaeftsfuehrung-von-Ausschuessen/AGS/pdf/AGS-publik-2013-7.pdf?__blob=publicationFile&v=2.

Byrne, J., Baugh, J. (2008). The significance of nanoparticles in particle-induced pulmonary fibrosis. McGill journal of medicine : MJM : an international forum for the advancement of medical sciences by students 11, 43-50.

Chen, R., Chen, C. (2016). Exposure scenarios in the workplace and risk assessment of carbon nanomaterials. In: Biomedical Applications and Toxicology of Carbon Nanomaterials (Ed), Chen, C.; Wang, H., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp. 515-534.

DFG, Deutsche Forschungsgemeinschaft. (2019). MAK- und BAT-Werte-Liste 2019. Ständige Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe. Mitteilung 55, WILEY-VCH Verlag GmbH, Weinheim. https://onlinelibrary.wiley.com/doi/book/10.1002/9783527826155.

Dixon, D., Herbert, R.A., Kissling, G.E., Brix, A.E., Miller, R.A., Maronpot, R.R. (2008). Summary of chemically induced pulmonary lesions in the National Toxicology Program (NTP) toxicology and carcinogenesis studies. Toxicologic Pathology 36, 428-439.

Donaldson, K., Borm, P.J., Oberdörster, G., Pinkerton, K.E., Stone, V., Tran, C.L. (2008). Concordance between *in vitro* and *in vivo* dosimetry in the proinflammatory effects of low-toxicity, low-solubility particles: the key role of the proximal alveolar region. Inhalation Toxicology 20, 53-62.

Dong, J., Ma, J., Tian, L., Inthavong, K., Tu, J. (2019). Inhalation Exposure Analysis of Lung-Inhalable Particles in an Approximate Rat Central Airway. International Journal of Environmental Research and Public Health 16.

ECHA, European Chemicals Agency. (2018). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Draft Version 3.0. November 2018, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

Ferron, G.A., Upadhyay, S., Zimmermann, R., Karg, E. (2013). Model of the deposition of aerosol particles in the respiratory tract of the rat. II. Hygroscopic particle deposition. Journal of Aerosol Medicine and Pulmonary Drug Delivery 26, 101-119.

FoBiG, Forschungs- und Beratungsinstitut Gefahrstoffe GmbH. (2011). Gutachten zur biologischen Plausibilität HEC und MPPD, Im Auftrag der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund/Berlin/Dresden. Bearbeitet von Dr. Fritz Kalberlah und Dr. Ulrike Schuhmacher-Wolz.

Fröhlich, E., Mercuri, A., Wu, S., Salar-Behzadi, S. (2016). Measurements of deposition, lung surface area and lung fluid for simulation of inhaled compounds. Front Pharmacol 7, 181.

Furi, P., Farkas, A., Madas, B.G., Hofmann, W., Winkler-Heil, R., Kudela, G., Balashazy, I. (2020). The degree of inhomogeneity of the absorbed cell nucleus doses in the bronchial region of the human respiratory tract. Radiation and Environmental Biophysics 59, 173-183.

Galle, P., Berry, J.P., Galle, C. (1992). Role of alveolar macrophages in precipitation of mineral elements inhaled as soluble aerosols. Environmental Health Perspectives 97, 145-147.

Ganguly, K., Carlander, U., Garessus, E.D.G., Fridén, M., Eriksson, U.G., Tehler, U., Johanson, G. (2019). Computational modeling of lung deposition of inhaled particles in chronic obstructive pulmonary disease (COPD) patients: identification of gaps in knowledge and data. Critical Reviews in Toxicology 49, 160-173.

Gebel, T., Foth, H., Damm, G., Freyberger, A., Kramer, P.-J., Lilienblum, W., Röhl, C., Schupp, T., Weiss, C., Wollin, K.-M., Hengstler, J.G. (2014). Manufactured nanomaterials: categorization and approaches to hazard assessment. Archives of Toxicology 88, 2191-2211.

Gehr, P., Bachofen, M., Weibel, E.R. (1978). The normal human lung: ultrastructure and morphometric estimation of diffusion capacity. Respiration Physiology 32, 121-140.

Geiser, M., Kreyling, W.G. (2010). Deposition and biokinetics of inhaled nanoparticles. Particle and Fibre Toxicology 7, 2-2.

Gregoratto, D., Bailey, M.R., Marsh, J.W. (2010). Modelling particle retention in the alveolar–interstitial region of the human lungs. Journal of Radiological Protection 30, 491-512.

Gregoratto, D., Bailey, M.R., Marsh, J.W. (2011). Particle clearance in the alveolar–interstitial region of the human lungs: model validation. Radiation Protection Dosimetry 144, 353-356.

Greim, H. (1997). Gesundheitsschädliche Arbeitsstoffe, Toxikologisch-arbeitsmedizinische Begründungen von MAK-Werten, Loseblattsammlung, 25. Lfg, DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag Weinheim.

Guha, S., Hariharan, P., Myers, M.R. (2014). Enhancement of ICRP's lung deposition model for pathogenic bioaerosolsl. Aerosol Science and Technology 48, 1226-1235.

Hartwig, A. (2012). Allgemeiner Staubgrenzwert (A-Fraktion) (Granuläre biobeständige Stäube (GBS)). In: T.-a.B.v.M.-W. Gesundheitsschädliche Arbeitsstoffe, Loseblattsammlung, 53. Lfg (Ed), DFG, Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag Weinheim

Hsieh, T.H., Yu, C.P., Oberdörster, G. (1999). Deposition and clearance models of Ni compounds in the mouse lung and comparisons with the rat models. Aerosol Science and Technology 31, 358-372.

ICRP, International Commission on Radiological Protection. (1994). Human Respiratory Tract Model for Radiological Protection. ICRP Publication 66, Pergamon Press.

Isaacs, K.K., Rosati, J.A., Martonen, T.B. (2005). Chapter VIII: Modeling Deposition of Inhaled Particles. In: D. Aerosols Handbook: Measurement, and Health Effects (Ed), Ruzer, L. S.
Harley, N. H., CRC Press LLC, Boca Raton, FL, pp. 113-155.

Jarabek, A.M. (1995). The application of dosimetry models to identify key processes and parameters for default dose-response assessment approaches. Toxicology Letters 79, 171-184.

Jarabek, A.M. (2016). Adaptive Dosimetry:
US EPA Approach to Address Evolving Science and Emerging Applications in Risk A ssessment, Granular Biopersistent
Dust (GBS) and Translational Toxicology: Deriving HECs/ Occupational Limit Values. Berlin, Germany, December 8 -9, 2016.

Jarabek, A.M., Asgharian, B., Miller, F.J. (2005). Dosimetric adjustments for interspecies extrapolation of inhaled poorly soluble particles (PSP). Inhalation Toxicology 17, 317-334.

Knust, J., Ochs, M., Gundersen, H.J.G., Nyengaard, J.R. (2009). Stereological estimates of alveolar number and size and capillary length and surface area in mice lungs. The Anatomical Record 292, 113-122.

Kolanjiyil, A.V., Kleinstreuer, C., Kleinstreuer, N.C., Pham, W., Sadikot, R.T. (2019). Mice-to-men comparison of inhaled drug-aerosol deposition and clearance. Respir Physiol Neurobiol 260, 82-94.

Koltermann-Jülly, J., Keller, J.G., Vennemann, A., Werle, K., Müller, P., Ma-Hock, L., Landsiedel, R., Wiemann, M., Wohlleben, W. (2018). Abiotic dissolution rates of 24 (nano)forms of 6 substances compared to macrophage-assisted dissolution and in vivo pulmonary clearance: Grouping by biodissolution and transformation. NanoImpact 12, 29-41.

Kreyling, W.G., Godleski, J.J., Kariya, S.T., Rose, R.M., Brain, J.D. (1990). *In vitro* dissolution of uniform cobalt oxide particles by human and canine alveolar macrophages. American Journal of Respiratory Cell and Molecular Biology 2, 413-422.

Krombach, F., Münzing, S., Allmeling, A.-M., Gerlach, J.T., Behr, J., Dörger, M. (1997). Cell size of alveolar macrophages: an interspecies comparison. Environmental Health Perspectives 105, Suppl. 5, 1261-1263.

Kuempel, E.D., Sweeney, L.M., Morris, J.B., Jarabek, A.M. (2015). Advances in inhalation dosimetry models and methods for occupational risk assessment and exposure limit derivation. Journal of Occupational and Environmental Hygiene 12, Suppl. 1, S18-40.

Kuempel, E.D., Tran, C.L., Smith, R.J., Bailer, A.J. (2001). A biomathematical model of particle clearance and retention in the lungs of coal miners. II. Evaluation of variability and uncertainty. Regulatory Toxicology and Pharmacology 34, 88-101.

Lejon, C. (2019). Lung Deposition Models for Exposure and Risk assessment. Report No. FOI-R--4753--SE FOI, Swedish Defence Research Agency, Stockholm, Sweden.

Lenfant, C. (2000). Discovery of endogenous surfactant and overview of its metabolism and actions. In: R. Notter (Ed), Lung Surfactants, Basic Science and Clinical Applications, New York, NY: Marcel Dekker, Inc., pp. 119-150.

Li, W., Pauluhn, J. (2018). Re-defining kinetic lung overload: time for new paradigms. Toxicology Letters 295, 212-219.

Löndahl, J., Möller, W., Pagels, J.H., Kreyling, W.G., Swietlicki, E., Schmid, O. (2014). Measurement techniques for respiratory tract deposition of airborne nanoparticles: a critical review. Journal of Aerosol Medicine and Pulmonary Drug Delivery 27, 229-254.

Mauderly, J.L. (1986). Respiration of F344 rats in nose-only inhalation exposure tubes. Journal of Applied Toxicology 6, 25-30.

Mauderly, J.L., Tesarek, J.E., Sifford, L.J., Sifford, L.J. (1979). Respiratory measurements of unsedated small laboratory mammals using nonbreathing valves. Laboratory Animal Science 29, 323-329.

Miller, F.J., Asgharian, B., Schroeter, J.D., Price, O. (2016). Improvements and additions to the Multiple Path Particle Dosimetry model. Journal of Aerosol Science 99, 14-26.

Miller, F.J., Asgharian, B., Schroeter, J.D., Price, O., Corley, R.A., Einstein, D.R., Jacob, R.E., Cox, T.C., Kabilan, S., Bentley, T. (2014). Respiratory tract lung geometry and dosimetry model for male Sprague-Dawley rats. Inhalation Toxicology 26, 524-544.

Miller, F.J., Kaczmar, S.W., Danzeisen, R., Moss, O.R. (2013). Estimating lung burdens based on individual particle density estimated from scanning electron microscopy and cascade impactor samples. Inhalation Toxicology 25, 813-827.

Morfeld, P., Bruch, J., Levy, L., Ngiewih, Y., Chaudhuri, I., Muranko, H.J., Myerson, R., McCunney, R.J. (2015). Translational toxicology in setting occupational exposure limits for dusts and hazard classification – a critical evaluation of a recent approach to translate dust overload findings from rats to humans. Particle and Fibre Toxicology 12, 3.

Morris, J., Asgharian, B., Kimbell, J. (2010). Upper airway dosimetry of gases, vapors, and particulate matter in rodents. Chapter 6. In: J.B. Morris, D.J. Shusterman (Eds), Toxicology of the Nose and Upper Airways, pp. 99-115.

Morrow, P.E. (1992). Dust overloading of the lungs: update and appraisal. Toxicology and Applied Pharmacology 113, 1-12.

Moss, O.R. (2010). Particle dosimetry in the nose and upper airways of humans. Chapter 8. In: J.B. Morris, D.J. Shusterman (Eds), Toxicology of the Nose and Upper Airways, pp. 122-128.

Muhle, H., Bellmann, B., Creutzenberg, O., Dasenbrock, C., Ernst, H., Kilpper, R., MacKenzie, J.C., Morrow, P., Mohr, U., Takenaka, S., Mermelstein, R. (1991). Pulmonary response to toner upon chronic inhalation exposure in rats. Fundamental and Applied Toxicology 17, 280-299.

NCRP, National Council on Radiation Protection and Measurements. (1997). Deposition, Retention and Dosimetry of Inhaled Radioactive Substances. NCRP Report No. 125, National Council on Radiation.

Nikula, K.J., Avila, K.J., Griffith, W.C., Mauderly, J.L. (1997). Lung tissue responses and sites of particle retention differ between rats and cynomolgus monkeys exposed chronically to diesel exhaust and coal dust. Fundamental and Applied Toxicology 37, 37-53.

Nikula, K.J., Vallyathan, V., Green, F.H., Hahn, F.F. (2001). Influence of exposure concentration or dose on the distribution of particulate material in rat and human lungs. Environmental Health Perspectives 109, 311-318.

NTP, National Toxicology Program. (1991). Toxicity Studies of Cobalt Sulfate Heptahydrate in F344/N Rats and B6C3F1 Mice (Inhalation Studies). Toxicity Report Series 5, U.S. Department of Health and Human Services; Public Health Service.

NTP, National Toxicology Program. (1998). Toxicology and Carcinogenesis Studies of Cobalt Sulfate Heptahydrate in F344/N Rats and B6C3F1Mice (Inhalation Studies). TR 471, U.S. Department of Health and Human Services; Public Health Service.

NTP, National Toxicology Program. (2014). Toxicology Studies of Cobalt Metal (CASRN 7440-48-4) in F344/N Rats and B6C3F1/N Mice and Toxicology and Carcinogenesis Studies of Cobalt Metal in F344/NTac Rats and B6C3F1/N Mice (Inhalation Studies). TR 581, U.S. Department of Health and Human Services. Public Health Service. https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr581_508.pdf.

Oberdörster, G. (1988). Lung clearance of inhaled insoluble and soluble particles. Journal of Aerosol Medicine 1, 289-330.

Oberdörster, G. (1995). Lung particle overload: implications for occupational exposures to particles. Regulatory Toxicology and Pharmacology 21, 123-135.

Oberdörster, G. (2010). Dosimetrische Extrapolation von Ergebnissen aus Rattenstudien zur Ableitung von Expositions-Standards, Vortrag Symposium Luftgrenzwerte, Berlin, 16.-17. September 2010.

OECD, Organisation for Economic Co-Operation and Development. (2018). Guidance Document on Inhalation Toxicity Studies. Series on Testing and Assessment No. 39. ENV/JM/MONO(2009)28/REV1, OECD Publishing, Paris. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2009)28/rev1&doclanguage=en.

OEHHA, Office of Environmental Health Hazard Assessment. (2018). Calculation of Rat Breathing Rate Based on Bodyweight Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. https://oehha.ca.gov/media/downloads/crnr/calcuratbreathingrate092818.pdf.

Oller, A.R., Oberdörster, G. (2010). Incorporation of particle size differences between animal studies and human workplace aerosols for deriving exposure limit values. Regulatory Toxicology and Pharmacology 57, 181-194.

Oller, A.R., Oberdörster, G. (2016). Incorporation of dosimetry in the derivation of reference concentrations for ambient or workplace air: A conceptual approach. Journal of Aerosol Science 99, 40-45.

Pauluhn, J. (2011a). Common mechanism-based study design for inhalation toxicity testing of poorly soluble (nano-/ultrafine-) particles: are new testing strategies and interpretation guidelines needed?

Pauluhn, J. (2011b). Poorly soluble particulates: searching for a unifying denominator of nanoparticles and fine particles for DNEL estimation. Toxicology 279, 176-188.

Paustenbach, D.J., Tvermoes, B.E., Unice, K.M., Finley, B.L., Kerger, B.D. (2013). A review of the health hazards posed by cobalt. Critical Reviews in Toxicology 43, 316-362.

Phalen, R.F., Mendez, L.B., Oldham, M.J. (2010). New developments in aerosol dosimetry. Inhalation Toxicology 22, Suppl. 2, 6-14.

Pott, F., Roller, M. (2001). The 19-Dust Carcinogenicity Study in Rats, Handout published at the 8th International Symposium 6-9 June 2001 Hannover Medical School Hannover Germany.

Price, O.T., Asgharian, B., Miller, F.J., Cassee, F.R., de Winter-Sorkina, R. (2002). Multiple Path Particle Dosimetry model (MPPD v1.0): A model for human and rat airway particle dosimetry. RIVM rapport 650010030 RIVM, Rijksinstituut voor Volksgezondheid en Milieu, Bilthoven, Netherlands.

Raabe, O.G., Al-Bayati, M.A., Teague, S.V., Rasolt, A. (1988). Regional deposition of inhaled monodisperse coarse and fine aerosol particles in small laboratory animals. Annals of Occupational Hygiene 32, Suppl. 1, 53-63.

Raabe, O.G., Yeah, H.-C., Schum, M.G., Phalen, R.F. (1976). Tracheobronchial Geometry: Human, Dog, Rat, Hamster - A Compilation of Selected Data from the Project Respiratory Tract Deposition Models. Lovelace Foundation Report LF-53, Lovelace Foundation for Medical Education and Research, Albuquerque, NM. https://digital.library.unt.edu/ark:/67531/metadc100754/.

Riediker, M., Zink, D., Kreyling, W., Oberdörster, G., Elder, A., Graham, U., Lynch, I., Duschl, A., Ichihara, G., Ichihara, S., Kobayashi, T., Hisanaga, N., Umezawa, M., Cheng, T.J., Handy, R., Gulumian, M., Tinkle, S., Cassee, F. (2019). Particle toxicology and health - where are we? Particle and Fibre Toxicology 16, 19.

RIVM, Rijksinstituut voor Volksgezondheid en Milieu. (1999). Multiple Path Particle Deposition Model, MPPDep version 1.11, Chemical Industry Institute of Toxicology Rijksinstituut voor Volksgezondheid en Milieu Bilthoven Netherlands.

Shang, Y., Dong, J., Inthavong, K., Tu, J. (2015). Comparative numerical modeling of inhaled micron-sized particle deposition in human and rat nasal cavities. Inhalation Toxicology 27, 694-705.

Snipes, M.B. (1989). Long-term retention and clearance of particles inhaled by mammalian species. CRC Critical Reviews in Toxicology 20, 175-211.

Snipes, M.B., McClellan, R.O., Mauderly, J.L., Wolff, R.K. (1989). Retention patterns for inhaled particles in the lung: comparisons between laboratory animals and humans for chronic exposures. Health Physics 57, Suppl. 1, 69-78.

Stebounova, L.V., Gonzalez-Pech, N.I., Peters, T.M., Grassian, V.H. (2018). Physicochemical properties of air discharge-generated manganese oxide nanoparticles: Comparison to welding fumes. Environmental Science. Nano 2018, 696-707.

Stone, K.C., Mercer, R.R., Gehr, P., Stockstill, B., Crapo, J.D. (1992). Allometric relationships of cell numbers and size in the mammalian lung. American Journal of Respiratory Cell and Molecular Biology 6, 235-243.

Stopford, W., Turner, J., Cappellini, D., Brock, T. (2003). Bioaccessibility testing of cobalt compounds. Journal of Environmental Monitoring 5, 675-680.

US EPA, Environmental Protection Agency. (1994). Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry, U.S. Environmental Protection Agency, Research Triangle Park NC. Office of Health and Environmental Assessment.

US EPA, Environmental Protection Agency. (2004). Air Quality Criteria for Particulate Matter. Vol. II of II. EPA/600/P-99/002bF, U.S. Environmental Protection Agency, Research Triangle Park, NC.

US EPA, Environmental Protection Agency. (2008). Draft TSD for Noncancer RELs, SRP2, June 2008, Appendix F: Estimating Human Equivalent Concentrations Using the U.S. EPA Default Approach, U.S. Environmental Protection Agency Washington DC.

Varghese, S.K., Gangamma, S. (2009). Particle deposition in human respiratory system: Deposition of concentrated hygroscopic aerosols. Inhalation Toxicology 21, 619-630.

Winkler-Heil, R., Ferron, G., Hofmann, W. (2014). Calculation of hygroscopic particle deposition in the human lung. Inhalation Toxicology 26, 193-206.

Yeh, H.-C., Cuddihy, R.G., Phalen, R.F., Chang, I.Y. (1996). Comparisons of calculated respiratory tract deposition of particles based on the proposed NCRP model and the new ICRP66 model. Aerosol Science and Technology 25, 134-140.

Yeh, H.C. (1980). Respiratory Tract Deposition Models. Final Report, Inhalation Toxicology Research Institute; Lovelace Biomedical & Environmental Research

Institute, Albuquerque, NM. Prepared for the National Institute of Environmental Health Sciences.

Yeh, H.C., Schum, G.M. (1980). Models of human lung airways and their application to inhaled particle deposition. Bulletin of Mathematical Biology 42, 461-480.

Zalkin, A., Ruben, H., Templeton, D.H. (1961). The Crystal Structure of Cobalt Sulfate Hexahydrate. In: R.L. UCRL-9802 (Ed), Lawrence, E. O., University of California. https://escholarship.org/uc/item/6mv5q6xk

Zhang, B., Qi, S., Yue, Y., Shen, J., Li, C., Qian, W., Wu, J. (2018). Particle disposition in the realistic airway tree models of subjects with tracheal bronchus and COPD. BioMed Research International 2018, 7428609.

List of Tables

List of Figures

REPORT 10:
Synthesis Report: Modelling of Distributions of Assessment Factors, Comparison with Current Methods and Discussion of Protection Goals

RESEARCH PROJECT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

Contents

Summary

This last report of the project “Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels” builds on the results of the previous project parts and discusses protection goals of the existing methodologies to derive occupational exposure limits (OELs) in the light of these results.

Using the empirical distributions for time, inter- and intraspecies extrapolation of toxicological data, parametric distributions were derived, which can be used in probabilistic modelling to describe the uncertainties of OELs. For time extrapolation, interspecies extrapolation (to be used in addition to allometric scaling) and intraspecies extrapolation (regarding differences in toxicodynamics) lognormal distributions were found, which fitted the empirical data well. For interindividual toxicokinetic differences the distribution of log GSD (geometric standard deviation) describing the toxicokinetic variability in the adult human population was used to establish a distribution.

$$\text{Distribution ratios}_{\text{Intra,TK i}} = 10^{\log_{10}\text{GSD}*z_{1-i}}$$

where $z_{1\text{-}i}$ is the z-Score of the normal distribution corresponding to the fraction of the population to be covered (calculations were performed for inclusion of 95% or 99% of the population).

By comparing with these distributions currently used assessment factors were discussed regarding the coverage (probability that the factor provides sufficient protection) achieved. Large differences in coverage provided were observed between different types of assessment factors (large coverage observed for subacute to chronic time extrapolation and interspecies extrapolation, lower coverage for subchronic to chronic time extrapolation and intraspecies extrapolation) as well as between the different frameworks. When the full set of assessment factors is compared with distributions combined by Monte Carlo simulation, the following sequence (with decreasing coverage) is observed:

BPR > RAC/REACH > AGS > ECETOC.

(BPR: Biocidal Products Regulation; RAC: Committee for Risk Assessment (in charge of deriving OELs at the EU level); REACH: ECHA guidance for deriving DNELs under REACH; AGS: Ausschuss für Gefahrstoffe, German OEL system; ECETOC: European Centre for Ecotoxicology and Toxicology of Chemicals).

The framework for assessing plant protection products is assumed to provide similar protection goals as BPR, although default values are not available for all extrapolation steps. The framework of the German MAK Commission (Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area of the Deutsche Forschungsgemeinschaft) also does also not provide default values for all of the steps but is expected to be situated between AGS and ECETOC. This sequence is mainly influenced by the size of the intraspecies extrapolation factor used in the different methodologies.

The parametric distributions obtained were also compared with those proposed for use in other probabilistic models. Further, the influence of the point of departure (POD: NOAEL/LOAEL or BMD) was discussed and exemplified with two example substances modelled probabilistically by Monte Carlo simulation. The position of the NOAEL or LOAEL relative to the BMD can vary substantially. As the uncertainty inherent to the NOAEL or LOAEL is not considered when using these PODs, benchmark dose modelling is the preferred way to derive the POD. It requires defining a benchmark response and, hence, also allows a clearer definition of the OEL for both quantal and continuous effect data. Probabilistic modelling allows to use the full uncertainty distribution of the BMD for the assessment.

Several recommendations were derived from the findings:

Recommendation 1:
All OEL derivation frameworks should clearly define their protection goals by stating:
- The fraction of the exposed population covered by the OEL
- The probability with which they intend to provide protection from adverse effects (as defined by the POD)

Recommendation 2:
Benchmark dose modelling should be used as the default procedure to derive a POD

Recommendation 3:
Probabilistic models should be further developed and used for benchmarking against deterministic methodologies to test them

Recommendation 4:
Increasing and improving the database on inter-individual variability in human inhalation studies might allow to establish route-specific distributions for intraspecies variability.

Abbreviations

AF	Assessment factor
AGS	Ausschuss für Gefahrstoffe
APROBA	Approximate probabilistic analysis
BAuA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
BMD	Benchmark dose
BMDL	Benchmark dose lower bound
BMDU	Benchmark dose upper bound
BMR	Benchmark response
BMDS	Benchmark dose software
BPR	Biocidal products regulation
CI	Confidence interval
DFG	Deutsche Forschungsgemeinschaft
DNEL	Derived no effect level
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
EFSA	European Food Safety Authority
GM	Geometric mean
GSD	Geometric standard deviation
IPCS	WHO's International Programme on Chemical Safety
LOAEC	Lowest observed adverse effect concentration
LOAEL	Lowest observed adverse effect level
MAK	Maximale Arbeitsplatzkonzentration

MC	Monte Carlo
MPPD	Multiple path particle dosimetry (model)
NOAEC	No observed adverse effect concentration
NOAEL	No observed adverse effect level
NTP	National Toxicology Program
OEL	Occupational exposure limit
PDF	Probability density function
POD	Point of departure
PPP	Plant protection products
PROAST	Dose-response modelling software by RIVM
QQ plots	quantile-quantile plots
RAC	Committee for Risk Assessment
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals,
RfD	Reference dose
RIVM	Dutch National Institute for Public Health and the Environment
SCOEL	Scientific Committee on Occupational Exposure Limits
SD	Standard deviation
SEM	Standard error of the mean
TD	Toxicodynamics
TK	Toxicokinetics
US EPA	Environmental Protection Agency in the US
WHO	World Health Organisation

1 Introduction

The overall objective of this project is to analyse and compare existing frameworks for deriving occupational exposure limits (OELs) and to discuss differences between the methods and protection goals of the OELs. The level of protection provided by currently used assessment factors to derive OEL from toxicological animal data is largely unknown and different sizes of assessment factors are a major reason for differences in OELs derived in the various frameworks.

In previous reports of this project, we established databases, derived empirical distributions from these databases (for the main assessment steps time, interspecies and intraspecies extrapolation) and discussed the results in the light of published proposals for such distributions. In the current report

- we use the empirical distributions to derive parametric distributions which can be used in probabilistic modelling approaches
- we compare the parametric distributions with currently used assessment factors and analyse the coverage (i.e., the probability that the OEL is protective enough) provided by the factors.

Further, the relevance and influence of different points of departure (POD) (i.e., BMDL or NOAEL) on the overall assessment is discussed. Here also the differences between continuous and quantal data and the consequences for the resulting OEL are considered.

With two example substances deterministic OELs derived according to existing methods are compared to the outcome of a probabilistic approach using the derived distributions. The influence of the POD as well as the coverage achieved by the deterministic OELs are discussed. All these comparisons are performed at two levels:

- for inclusion of 95% of the exposed population (i.e., an incidence level of 5%)
- for inclusion of 99% of the exposed population (i.e., an incidence level of 1%).

These choices are for illustration only and are not intended to pre-empt regulatory decisions on protection goals.

With these comparisons the protection goals provided by the various methodologies are analysed and discussed. This discussion also addresses similarities and differences in the evaluation of systemic versus local effects in the respiratory tract. Finally, we present recommendations how to increase transparency and harmonisation of the existing methodologies.

2 Distributions for extrapolation steps

2.1 Methods for parametrization of the empirical distributions

2.1.1 Examination of lognormality

Due to the shape of the empirically determined distributions in the reports on “Intraspecies extrapolation”, “Interspecies extrapolation” and “Time extrapolation” and the literature discussed therein (which all used lognormal distributions to model the data), it is reasonable to assume that the empirically determined distributions can be adequately described by lognormal distributions. To evaluate this assumption, quantile-quantile plots (QQ plots) of all distributions were prepared. In a QQ plot, the empirical quantiles of the distribution are plotted against the theoretical quantiles of the distribution function to be tested. If the empirical and theoretical quantiles are identical, the plotted points lie on a perfect straight line. In the case of a theoretical normal distribution, the intersection of this line with the y axis maps to the mean and its slope to the standard deviation of the distribution.

In **Figure 2-1** exemplary QQ plots with the empirical distribution for the time extrapolation subacute-chronic from NTP studies and a normal or a lognormal distribution are shown. The empirical data fits reasonably well to the lognormal distribution (since the plotted points lie on a straight line), while the theoretical quantiles of the normal distribution clearly deviate from the empirical distribution.

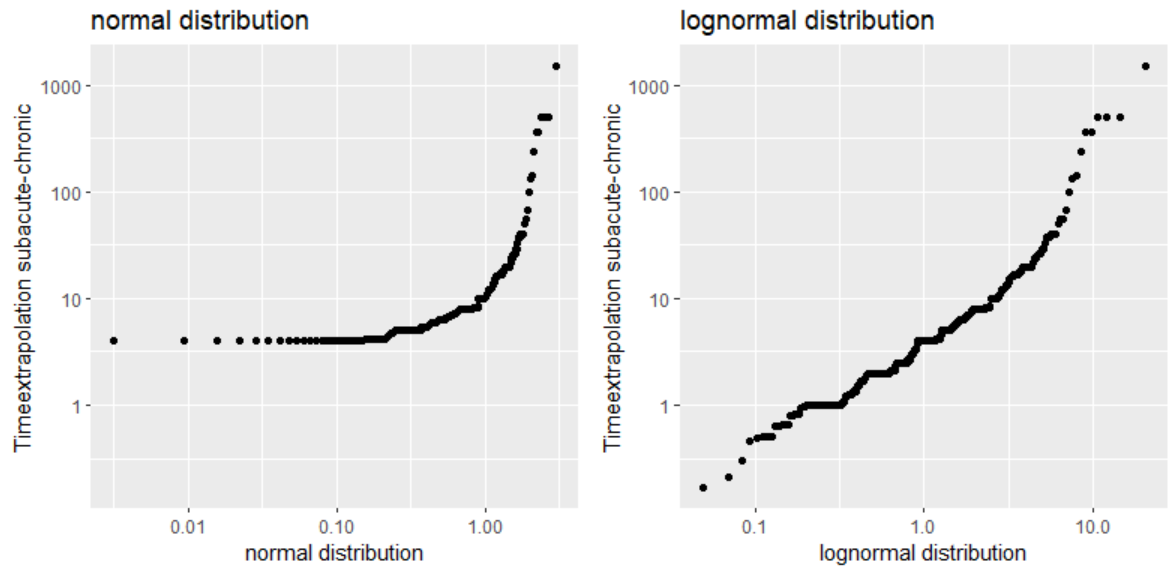


Figure 2-1 QQ-Plots of the empirical subacute-chronic extrapolation data from NTP studies against a theoretical normal and a lognormal distribution

However, “step-shaped” deviations from the theoretical distribution can be observed in the QQ plot. This was also seen and explained in the respective reports on the individual extrapolation steps and is a result of using rounded concentrations (usually to halves or thirds of log10 units) for dosing. This leads to an overrepresentation of fractions of two small integer values (e.g., 1/3, 1/5, 3/10, etc.) in the empirical ratio distribution. As this feature of the distribution is artificially introduced by experimental design choices, it is no reason to deviate from the assumption that the data can be adequately represented by a lognormal distribution. This feature is also the reason why statistical tests for (log)normality like the Lilliefors test reject the test hypothesis when applied to our (log transformed) data. We therefore did not consider statistical tests to examine the assumption of lognormality. It is an interesting feature of QQ plots to visualize the character of a deviation between empirical and theoretical distributions. The left-hand panel in **Figure 2-1** does suggest a search for an alternative theoretical distribution, while the right one does not. Such information is not easily obtained from numerical goodness-of-fit measures.

2.1.2 Derivation of μ and σ

In order to determine the parameters μ (location parameter, corresponding to the expected value on the log scale) and σ (shape parameter, corresponding to the standard deviation on the log scale) of a lognormal distribution for our data, the function ‘get.lnorm.par()’ from the R package ‘rriskDistributions’[1] v 2.1.2 (Belgorodski et al. 2017) was used, which uses R’s built-in optimization function with the “L-BFGS-B” method for fitting the parameters of a probability function to the empirical data. The function was called with the sorted empirical quantiles, the corresponding probabilities ($p_i = \frac{i-0.5}{n}$, where *n* is the number of empirical data points and $i \in \mathbb{N}^+, i \leq n$) and the default values for other parameters. Estimation results from the rriskDistributions package are preferred to solutions from the simple QQ plot-based regression, because they provide a more precise iterative solution.

An example of a parametrized lognormal distribution using the described fitting procedure on the data from above (ratios for the time extrapolation subacute-chronic) is shown in **Figure 2-2**. The density of the empirical data in this figure, as well as the following figures in this report, is estimated using the ggplot2 wrapper for the standard function for kernel density estimation in R (‘density()’) or its ggplot2 wrapper (‘stat_density()’) with standard parameters except ‘adjust = 1.5’. An R Markdown script of the fitting process, exemplary shown for the subacute-chronic data, is attached in Annex 1.

[1] The ‘rriskDistributions’ package was developed by a group of statisticians led by the German Federal Institute for Risk Assessment. It is part of the ‘rrisk’ package for risk modelling and reporting (hence the name). While ‘rriskDistributions’ is listed in the curated CRAN package repository, at the time of writing ‘rrisk’ is not and needs to be downloaded from the project webpage (https://www.bfr.bund.de/de/rrisk__risikomodellierung_und_automatische_berichterstellung_in_r-52158.html)

In order to be consistent with seminal publications on this topic, we use ‘log’ to describe the logarithm to base 10, while the natural logarithm is explicitly written as ‘ln’ in this report. Accordingly, the variable ‘log GSD’, which is often used in literature to describe the spread of distributions for intraspecies variability is used by us as such and refers to $\log_{10}$ GSD.

Although $\log_{10}$ transformations are easier to convert to the normal scale, we intentionally use the natural logarithm in figures showing lognormal distributions in order to stay consistent with the definition of the parameters μ and σ.

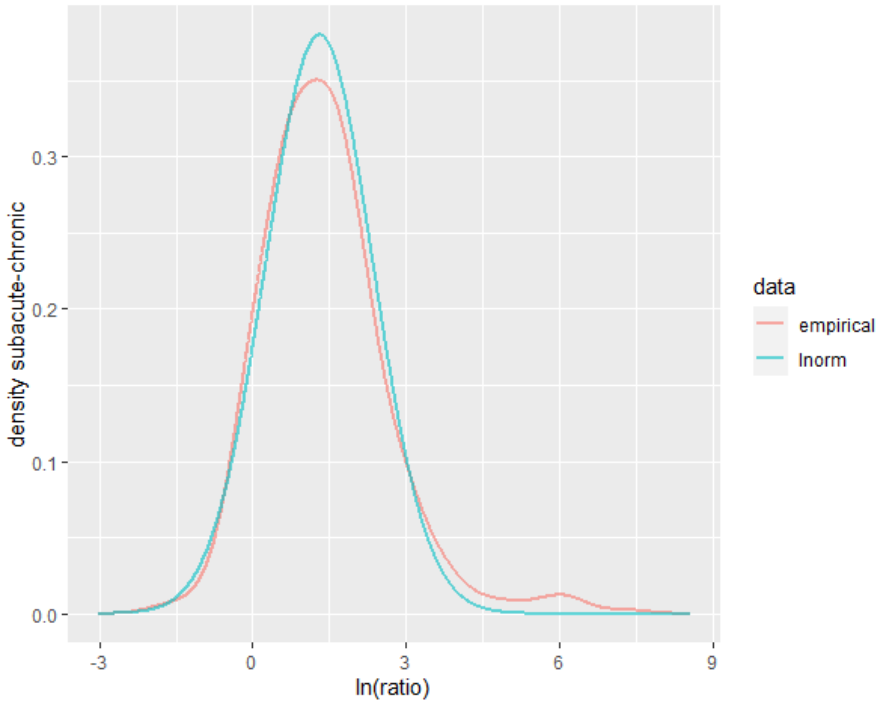


Figure 2-2 Density of the empirical distribution (obtained by kernel density estimation) of the subacute-chronic time extrapolation ratios (red) and the corresponding parametrized lognormal distribution (blue)

2.1.3 Monte Carlo simulations

Several times in this report a distribution is simulated by the Monte Carlo (MC) method. MC is a mathematical approach in statistics for deriving the distribution of an estimated quantity by taking a large number of repeated random samples from an expression with one or more random variables. For example, if a certain risk can be modelled by the multiplication of two distributions with known parameters, it may be difficult to find formally the induced probability function and its parameters that correspond to the risk distribution. Instead, a sufficiently large number of simulated samples from the two distributions can be multiplied with each other to generate a distribution corresponding to the investigated risk (for a more detailed explanation see the Annex 2 of our report “Probabilistic hazard assessment”). In this report, usually $1x10^7$ samples are used for

MC simulations, but plots were generated with a reduced sample number for the sake of performance.

Combining distributions by MC simulations again results in a distribution (for example the distribution resulting from combining the individual extrapolation steps time, inter- and intraspecies extrapolation, see section 3.5). The percentiles of such a distribution provide information on the probability. The coverage or probability achieved by an assessment factor can be read directly from the percentile value. For example, if an assessment factor has a value which corresponds to the 60th percentile of the respective distribution for this extrapolation step, then there is a 60% probability that the factor is sufficient to cover the uncertainties of this extrapolation step. Or in other words, the OEL resulting from application of this factor would be conservative enough for 60 out of 100 newly assessed substances.

EFSA is hosting an R Shiny application with the name 'MonteCarlo'[2], which provides a largely intuitive graphical user interface geared towards probabilistic risk assessment. The EFSA MonteCarlo application was used for the two examples in section 5, the remaining simulations were run in R using base functions[3].

2.2 Stratifications

During the analysis of empirical databases for the various extrapolation steps special emphasis was given to investigate various potential influencing parameters:

- Exposure route,
- Experimental species,
- Sex,
- Endpoint (body weight, systemic or local effects),
- Target organ (liver, kidney, others) or
- Structural similarity of substances.

Time extrapolation

Slight differences were observed for some parameters. Time extrapolation ratios obtained for inhalation exposure were slightly lower than those for oral exposure. Ratios for mice yielded slightly higher values than for rats for the comparisons subacute to subchronic and to chronic and for subchronic to chronic. But neither the differences according to exposure route nor to species did reach statistical significance.

No relevant differences were found with respect to sex or target organ. The chemical classes used to check for the influence of the chemical nature showed high variability and were unlikely to reveal meaningful differences.

In the case of the NTP data from inhalation studies it was possible to compare data from subchronic and chronic studies in relation to the type of adverse endpoint

[2] https://r4eu.efsa.europa.eu/ (requires registration)

[3] For each distribution in the simulation, 1x10^7 random samples were generated and combined using multiplication, division, etc.

(systemic versus local effects in the respiratory tract). Ratios derived for local effects were similar to those obtained for systemic effects. So, a concentration dependency (requiring lower assessment factors for time extrapolation) as assumed by ECETOC in their proposals for assessment factors to derive DNELs was not confirmed in our analysis (ECETOC 2003, 2010). This was similarly found by Schröder et al. (2015) where higher ratios for local effects (eye and respiratory tract) are reported for all comparisons (subacute to chronic, subacute to subchronic, subchronic to chronic).

In conclusion, no useful stratification was found in our analysis of data for time extrapolation. Both for subacute to chronic and subchronic to chronic extrapolation the same distributions are proposed to be used for oral and inhalation data and for systemic and local effects after inhalation (for details see our report on “Time extrapolation”).

Interspecies extrapolation

No meaningful difference was found for the parameters study duration, sex, or target organ (liver or kidney versus other target organs), when comparing rats and mice NTP studies. Regarding chemical nature, data from oral studies did not show a difference between the two evaluated groups (alkylated aromatics, metal compounds) when compared to the remaining substances. Inhalation data of alkylated aromatics point to a higher sensitivity of mice (at the 75th percentile only), whereas in the case of metal compounds, rats appeared to be more sensitive than mice. Overall, these differences are more likely the consequence of the small datasets and a high variability than real substance-group specific differences.

Slightly lower geometric means of the rats/mice ratios were found with NTP inhalation studies for local as well as systemic effects compared to the endpoint body weight. However, these differences vanished at the 75th percentile level. No differences were seen between local and systemic effects (see the separate report on “Interspecies extrapolation”).

Intraspecies extrapolation

In the analysis of human toxicokinetic data higher log GSD values (indicating higher variability in the human population under study) were found for oral compared to inhalation data. These differences were statistically significant when comparing GM values, but not at the 75th percentile level.

Similar differences were found when comparing pharmaceutical substances with industrial chemicals. However, the two parameters exposure route and substance class were not independent, as the large majority of inhalation data came from studies with industrial chemicals. These findings point to a lower intraspecies variability in case of inhalation exposure and/or industrial chemicals compared to oral exposure and/or pharmaceuticals. However, the opposite effect was found in a set of substances, for which information on effects after exposure to various exposure levels were available. These datasets describe variability in human groups at the level of toxicokinetic and -dynamic differences combined. This set of data is limited with regard to number and

preciseness of quantification of variability but point to the opposite direction: for effects caused by industrial chemicals in studies with inhalation exposure, a higher variability was observed as with oral data. Eleven of these twelve data had local effects in the respiratory tract as an endpoint.

Because of these diverging results we did not stratify the distribution for toxicokinetic differences according to route. However, we recommend to increase the database of toxicokinetic studies to verify whether there is a route-specific difference in the variability of human responses towards toxic chemicals.

No other stratification is possible, as the parameters route and substance group were highly interrelated (studies with industrial chemicals were mostly inhalation studies, pharmaceuticals mainly administered orally) (see the separate report on “Intraspecies extrapolation”).

Conclusions

In conclusion, we did not find differences in the datasets for time extrapolation, inter- and intraspecies extrapolation, which would have justified to stratify according to one of the parameters exposure route, species, sex, endpoint, target organ, or chemical class. The most important finding was a difference in the human variability found in oral versus inhalation studies, which warrants further examination.

Takeshita et al. (2014) proposed to use a different time extrapolation factor (subacute – subchronic) for adverse effects on blood than for other effects. For other target organs (liver, kidney, body weight) they did not find significant differences to the whole dataset. However, the authors’ conclusions are based on a very small set of data (9 for effects on blood) with a large variability and the significant difference for effects on blood might well be by chance. Our analysis of time extrapolation is performed on a much larger number of studies with different exposure durations and could not find consistent differences.

Therefore, in the following we derive distributions for each extrapolation step without further stratification, based on the empirical distributions found and described in the previous reports.

2.3 Time extrapolation

2.3.1 Introduction

In the project report on “Time extrapolation”, empirical distributions were derived to extrapolate from dose descriptors of shorter study durations to longer study durations. For each time extrapolation such a distribution was derived from NTP study data as well as from studies registered under REACH. It was concluded that the data derived from the NTP studies is most appropriate to use as the distribution to be used to model time extrapolation in a probabilistic hazard assessment.

For regulatory purposes, the time extrapolation steps subacute – chronic and subchronic – chronic are relevant. The respective se empirical distributions from the report on “Time extrapolation” are again shown in **Table 2-1**.

Table 2-1 Ratio distributions for the time extrapolations relevant for a probabilistic hazard assessment, derived from an evaluation of NTP study data

Comparison	Mean	SD	GM	GSD	5%	Median	75%	95%	n
subacute/ chronic	17.08	91.30	4.11	3.40	0.98	4.00	7.91	30.31	396
subchronic/ chronic	6.81	27.79	2.93	3.04	0.50	2.67	5.00	18.94	1218

2.3.2 Subacute – chronic

The distribution of factors for subacute-chronic extrapolation as shown in **Table 2-1** is well described by a lognormal distribution with the parameters **μ = 1.31 and σ = 1.05** (**Figure 2-1** and **Figure 2-2**). Moments and percentiles of the distribution with these parameters are shown in **Table 2-2**.

Table 2-2 Moments and characteristic percentiles of the parametrized distribution for subacute - chronic extrapolation

Extrapolation	GM	GSD	5%	Median	75%	95%
subacute/ chronic	3.71	2.86	0.66	3.71	7.52	20.85

2.3.3 Subchronic – chronic

The distribution of factors for subchronic-chronic extrapolation is well described by a lognormal distribution with the parameters **μ = 1.04 and σ = 0.99** (**Figure 2-3**). Moments and percentiles of the distribution with these parameters are shown in **Table 2-3**.

Table 2-3 Moments and characteristic percentiles of the parametrized distribution for subchronic – chronic extrapolation

Extrapolation	GM	GSD	5%	Median	75%	95%
subchronic/ chronic	2.83	2.70	0.55	2.83	5.53	14.49

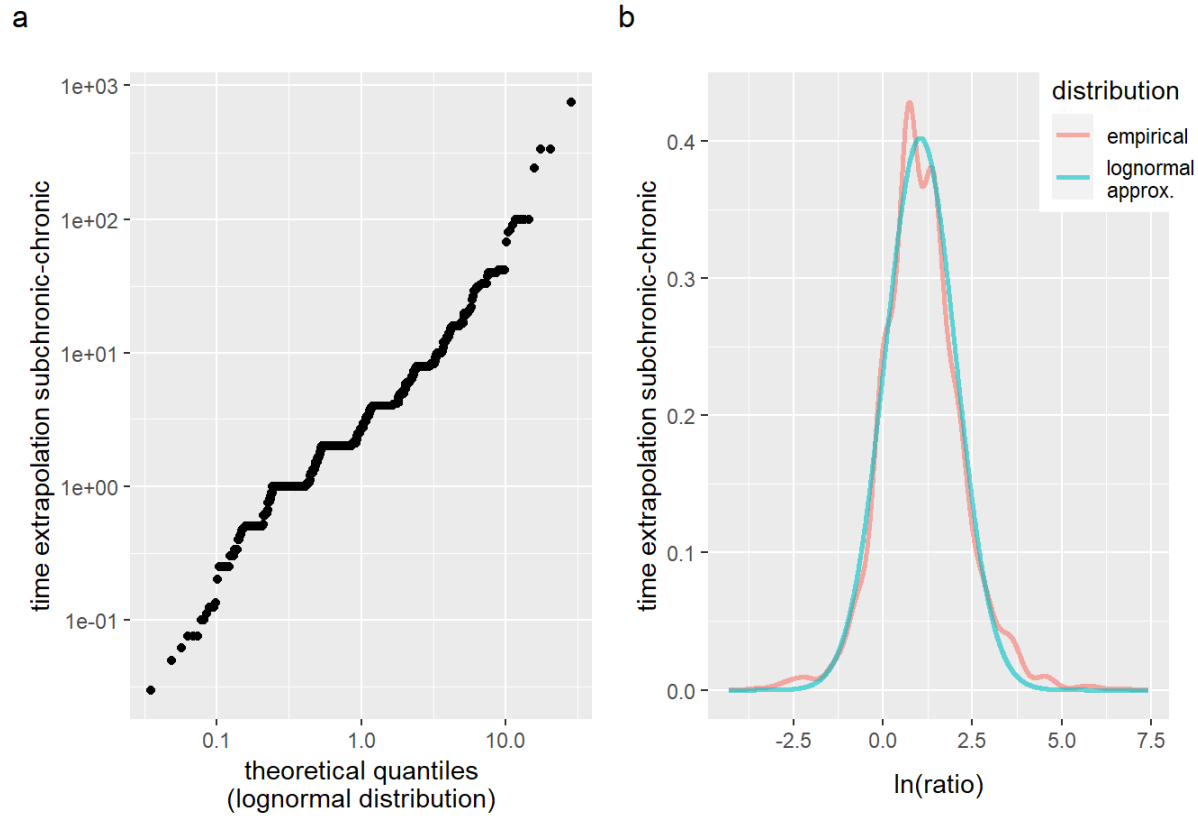


Figure 2-3 Lognormal approximation of the empirical distribution of the subchronic-chronic time extrapolation data. QQ-Plot of the empirical quantiles against the theoretical quantiles of a lognormal distribution (a). Density of the empirical distribution (b, red) and the corresponding parametrized lognormal distribution (b, blue).

2.4 Interspecies extrapolation

2.4.1 Introduction

In the report on “Interspecies extrapolation”, we recommended that after application of a correction factor according to caloric demand scaling, the remaining uncertainty due to substance-to-substance variability should be reflected by a distribution with a GM = 1. We derived data for interspecies extrapolation factors from NTP and REACH study data based on NOAEL ratios, of which the NTP data provides the better quality and should be used for modelling (**Table 2-4**).

We also discussed that additional variability introduced by using NOAEL ratios should be taken into account, if possible. To this end, and to characterise the magnitude of the variability, we compare the results of Bokkers and Slob (2007) to the distribution derived from our data evaluation. Bokkers and Slob (2007) derived a distribution for oral repeated dose studies based on BMD ratios, which in principle is better suited to characterize the uncertainty for interspecies extrapolation, because it avoids additional

uncertainty introduced by using ratios of NOAELs. They also derived distributions based on NOAEL ratios and compared these to the distributions of BMD ratios (**Table 2-4**). We believe that our data evaluation provides a more robust data basis to model interspecies differences than the evaluation in Bokkers and Slob (2007). However, we argue that the difference between the GSD from NOAEL ratios and the GSD from BMD ratios (which is further corrected by Bokkers and Slob (2007) for the error from individual BMD calculations) is an adequate measure for the additional spread introduced by using NOAEL ratios compared to true BMD ratios (Bokkers and Slob 2007). Further, as our data evaluation is based on NOAEL ratios, we can apply this correction of the GSD to our data to derive a distribution that reflects the ratios of the true (presumably error free) effect levels. This correction is performed in section 2.4.3.

Table 2-4 Overview of interspecies comparison based on NTP data

Species (species1/ species2)	**Route**	**Study source**	**GM (95% CI)**	**Expected value***	**GSD**
rat/ mouse	oral	own evaluation of NTP data	0.40 (0.37-0.44)	0.59	3.78
rat/ mouse	inhalation	own evaluation of NTP data	0.96 (0.84-1.10)	1.00	3.61
mouse/ rat	oral	Bokkers and Slob (2007) using NOAEL ratios**	1.88** (1.56 – 2.28)	1.69***	3.08
mouse/ rat	oral	Bokkers and Slob (2007) using BMD ratios without GSD correction**	1.86** (1.63 – 2.12)	1.69***	2.19
mouse/ rat	oral	Bokkers and Slob (2007) using BMD ratios with GSD correction**	1.86** (1.63 – 2.12)	1.69***	1.97

* Expected values according to caloric demand scaling: body weight species $2^{0.25}$/body weight species $1^{0.25}$, see report “Interspecies extrapolation”, Section 4.3
** For analysing the influence of using BMD versus NOAEL ratios here only the data was used were both BMD and NOAEL are available (reduced dataset in column 6 of Table 2 in Bokkers and Slob (2007)
*** In this evaluation the values from mice studies are divided by the values from rats and the expected ratio values are the inverse of our evaluation

2.4.2 Distribution from our own data evaluation

We determined separate distributions of factors from rat/mouse comparisons for the oral and inhalation route and did not perform a correction according to caloric demand scaling. Accordingly, the GM of the distributions for the two exposure routes deviate (oral 0.40, inhalation 0.96). As explained above, the correction according to caloric demand scaling should not be incorporated into the uncertainty distribution but rather accounted for by application as a separate factor. To this end, the distributions for the two routes are normalised to a GM of 1 by dividing all values in the distribution by the respective GM and the parametrized distribution is derived from the combined

normalised datapoints. The resulting distribution is described by a lognormal distribution with the parameters **μ = 0.02 and σ = 1.17** (**Figure 2-4**).

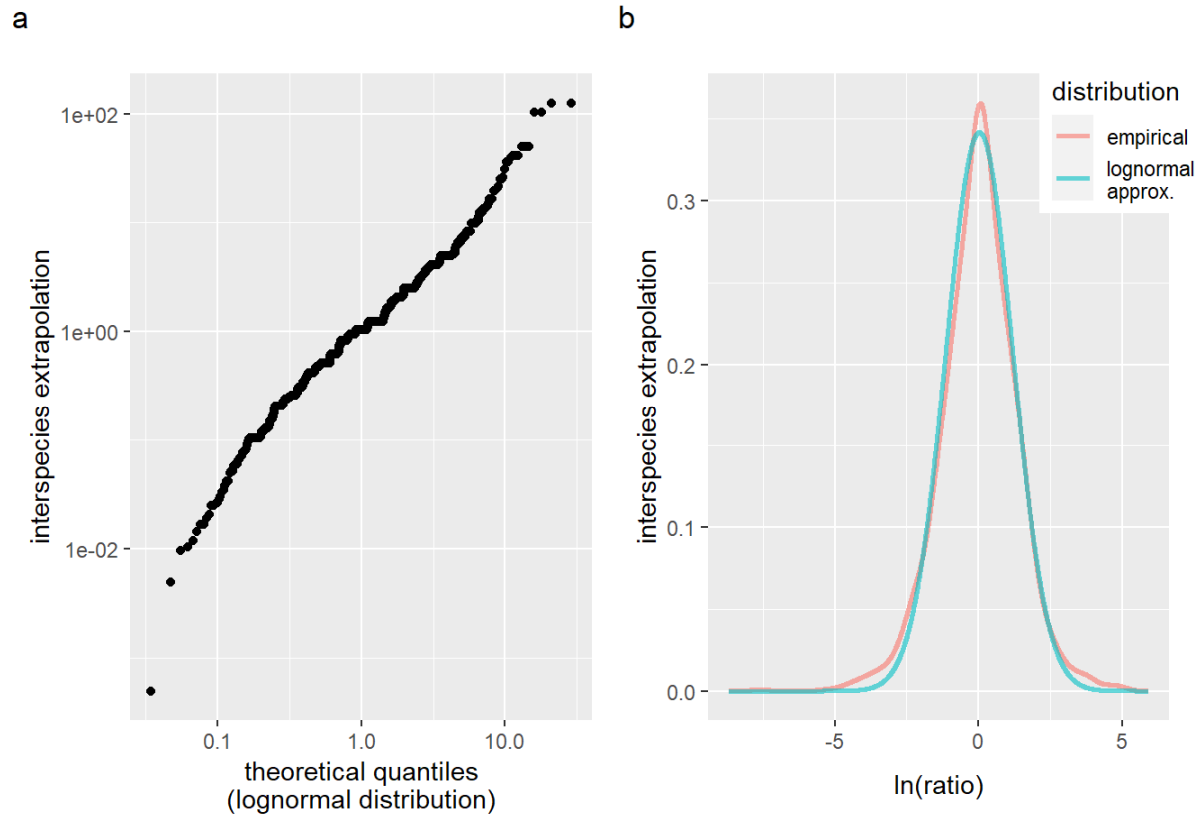


Figure 2-4 Lognormal approximation of the empirical distribution of the interspecies extrapolation data. QQ-Plot of the empirical quantiles against the theoretical quantiles of a lognormal distribution (a). Density of the empirical distribution (b, red) and the corresponding parametrized lognormal distribution (b, blue).

2.4.3 Correction for the errors introduced by using NOAEL ratios

As introduced above, the distribution derived from our data evaluation is further corrected for additional variance introduced by using NOAEL ratios, because the NOAEL carries inherent uncertainty. The evaluation by Bokkers and Slob (2007) provides data for this correction as they derived ratios based on BMD as well as on NOAEL for the same study selection (column 6 of Table 2 in Bokkers and Slob (2007)). The variance of the interspecies ratios derived from NOAEL ratios ($var_{NOAEL,\ uncorrected}$) is assumed to be composed of the variances for the ratios derived from the "true" effect levels ($var_{BMD,\ corrected}$) and the variance from the uncertainty of the NOAEL ($var_{NOAEL\text{-}error}$):

$$var_{NOAEL-error} = {\sigma^2}_{NOAEL-error} = var_{NOAEL,uncorrected} - var_{BMD,\ corrected}$$

where var and σ refer to the values on the logarithmic scale, hence the variance is subtracted. As var = σ² and σ = ln GSD (assuming lognormality), this results in:

$$\sigma_{NOAEL-error} = \sqrt{\left(\ln GSD_{NOAEL,uncorrected}\right)^2 - \left(\ln GSD_{BMD,\ corrected}\right)^2}$$

The values for $GSD_{NOAEL,uncorrected}$ and $GSD_{BMD,\ corrected}$ from Bokkers and Slob (2007) are 3.08 and 1.97, respectively (see **Table 2-4**). Using this relationship, $\sigma_{NOAEL-error}$ is 0.90.

Correspondingly, σ from fitting a lognormal distribution to our data evaluation in section 2.4.2 is corrected:

$$\sigma_{corrected} = \sqrt{\left(\sigma_{NOAEL,uncorrected}\right)^2 - \left(\sigma_{NOAEL-error}\right)^2}$$

Which yields a $\sigma_{corrected}$ of 0.75.

Taken together, the recommended distribution for interspecies differences after correction for the error introduced by using NOAELs is a lognormal distribution with the parameters **μ = 0.02 and σ = 0.75.** Moments and percentiles of the distribution with these parameters are shown in **Table 2-5**.

Table 2-5 Moments and characteristic percentiles of the parametrized distributions for interspecies extrapolation

Extrapolation	GM	GSD	5%	Median	75%	95%
Interspecies	1.02	2.11	0.3	1.02	1.69	3.49

2.5 Intraspecies extrapolation

An assessment factor to cover inter-individual differences in sensitivity to the substances humans may be exposed to needs to consider

- the fraction of the population which shall not show an effect, i.e. be protected by the assessment factor
- the uncertainty associated with exposure to a given, unknown substance.

Hence, in order to derive a distribution for describing uncertainty in intraspecies extrapolation, a decision needs to be taken on the percentage of the target population to be protected by the approach (or, the other way round, the remaining incidence – the percentage of the individuals with higher susceptibility needs to be determined).

Without prejudice to conclusions by regulatory bodies, for illustrative purposes we chose two levels:

- 5% incidence, i.e. 95% of the target population are covered
- 1% incidence, i.e. 99% of the target population are covered.

Uncertainty distributions for both target levels are derived in the following.

In practice, data evaluations to derive such distributions usually choose one of the following two approaches. For a number of substances, data on the inter-individual sensitivity is analysed. This inter-individual sensitivity is assumed to follow a lognormal distribution and a log GSD value may be derived for each substance (for more details on log GSD see Annex 2 - ”the concept of log GSD” in the report on “Intraspecies extrapolation”). In order to derive factors from the log GSD values, they need to be transformed according to the chosen incidence level in the population. The distribution of the transformed log GSD values can then be used for modelling the assessment factor in a probabilistic assessment (approach 1).

Alternatively, data on the inter-individual sensitivity may be derived from the studies as dose ratios e.g. the dose with 95% incidence divided by the dose with 50% incidence. In this case, the derived ratios represent assessment factors applicable for the respective target incidence rates and the distribution of these factors may be used for the probabilistic assessment (approach 2).

For deriving parametrized distributions describing the toxicokinetic and the toxicodynamic part to intraspecies uncertainty, we use both approaches, according to the way the data was derived (approach 1 for toxicokinetic variability, approach 2 for toxicodynamic variability).

2.5.1 Toxicokinetic variability

For toxicokinetic variability, our evaluation of log GSD values for the inter-individual toxicokinetic differences is used (**Table 2-6**).

Table 2-6 Parameters of the distribution of the log GSD values from all evaluated studies for toxicokinetics intraspecies extrapolation

Mean	SD	GM	GSD	5%	Median	75%	95%	n
0.166	0.094	0.141	1.83	0.049	0.146	0.22	0.355	68

This distribution of log GSD values is well described by a lognormal distribution with the parameters **μ = -1.93 and σ = 0.61** (**Figure 2-5**).

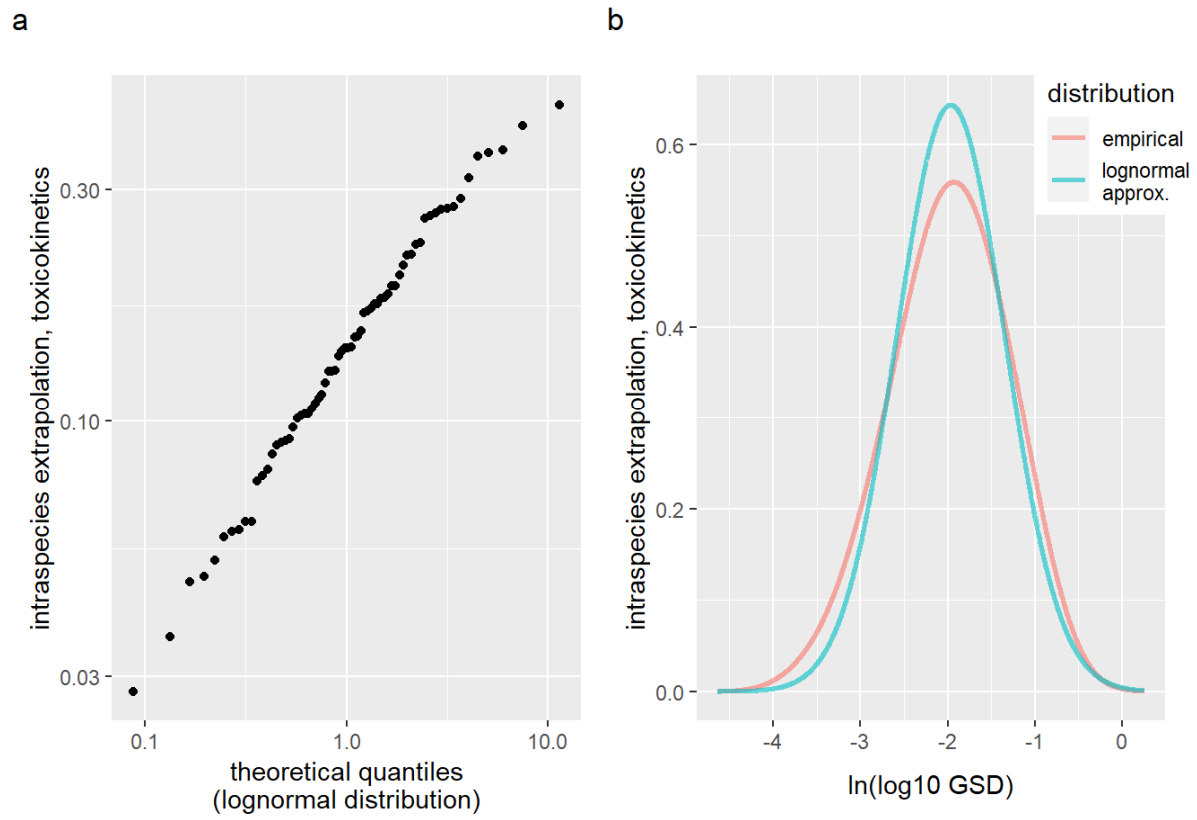


Figure 2-5 Lognormal approximation of the empirical distribution of log10 GSD values for interindividual variability in toxicokinetics. QQ-Plot of the empirical quantiles against the theoretical quantiles of a lognormal distribution (a). Density of the empirical distribution (b, red) and the corresponding parametrized lognormal distribution (b, blue). Please note that the natural logarithm of the log10 GSD values is plotted

In order to derive a distribution of assessment factors, the log10 GSD values are transformed by the relationship in Equation 1 corresponding to equation (4-3) in the WHO/IPCS report on uncertainties in hazard assessment (WHO 2014) with incidences of 5% and 1%, i.e. $z_{0.95}$ = 1.6449 and $z_{0.99}$ = 2.3263.

$$\begin{aligned} \mathrm{Ratio}_{\mathrm{Intra,\,i}} &= \text{Factor covering } (1-i) \text{ of the population} \\ &= 10^{\log_{10}\mathrm{GSD}*z_{1-i}} \end{aligned} \qquad \text{(Equation 1)}$$

where $z_{1\text{-}i}$ is the z-Score[4] of the normal distribution corresponding to the fraction of the population to be covered.

[4] The z-score is a measure of how many standard deviations below or above the mean a value is: a z-score of 1 is 1 standard deviation above the mean (representing the 84th percentile of the distribution), a z-score of 2 is 2 standard deviations above the mean (at the 97.5th percentile of the distribution); the chosen z-scores of 1.6449 and 2.3263 represent the 95th and 99th percentile of the normal distribution

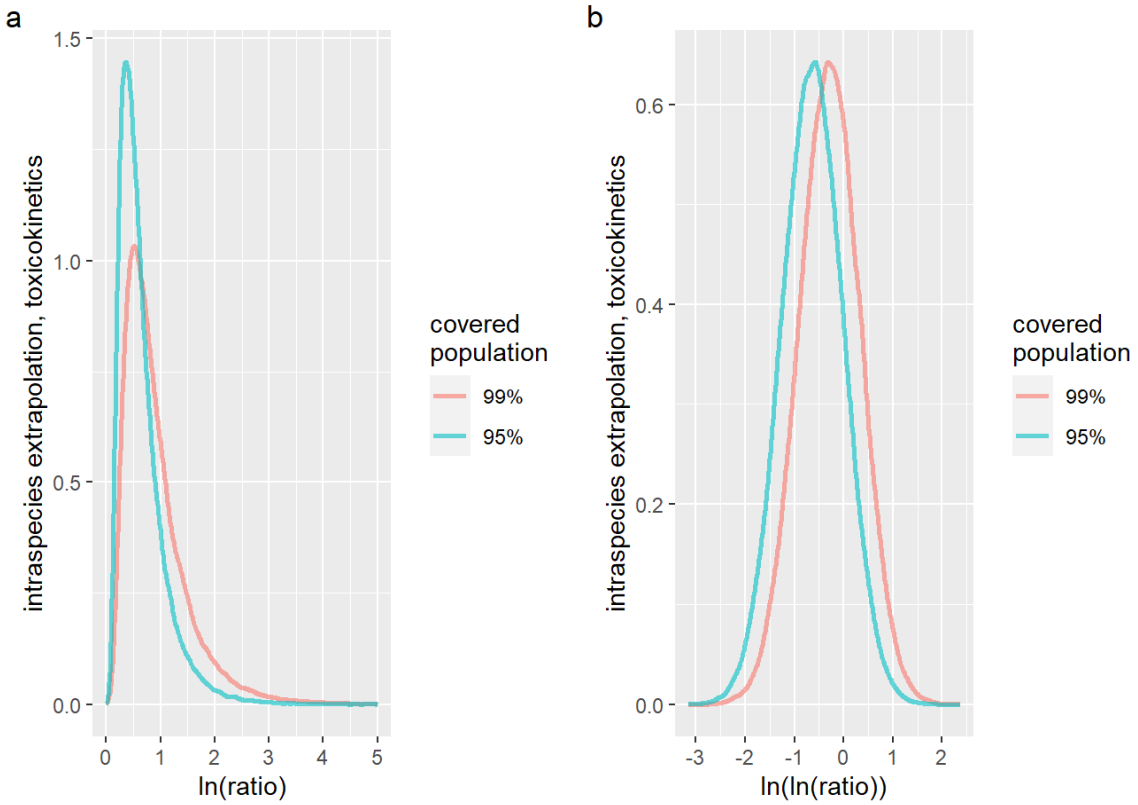


Figure 2-6 Distribution of the AF for interindividual variability in toxicokinetics, derived from transformation of the parametrised log10 GSD distributions. Densities of the distributions corresponding to 1% incidence (red) and 5% incidence (blue) are plotted after logarithmic transformation (a) and after an additional logarithmic transformation (b).

The resulting distributions (**Figure 2-6**) do not follow the characteristics of basic probability functions which are implemented in ready-to-use probabilistic tools. However, these distributions arise as strictly monotone transformations of standard distributions. Therefore, it is not necessary to parametrise the transformed distributions themselves and Equation 1 is used instead to model the assessment factors using samples from the parametrised lognormal distribution for $\log_{10}$ GSD as determined above.

2.5.2 Toxicodynamic variability

To model the toxicodynamic uncertainty, the data from Abdo et al. (2015) is used. Abdo et al. (2015) determined distributions for inter-individual variability from in vitro toxicity screenings of many substances on a large pool of lymphoblastoid cell lines derived from over 1000 human individuals. Factors that correspond to incidences of 5% and 1% within the screened population were derived for each tested chemical. They further corrected these factors for sampling variability. These two distributions of corrected

factors are used to derive parameters describing the toxicodynamic part of inter-individual variability.

2.5.2.1 Distribution corresponding to 5% incidence

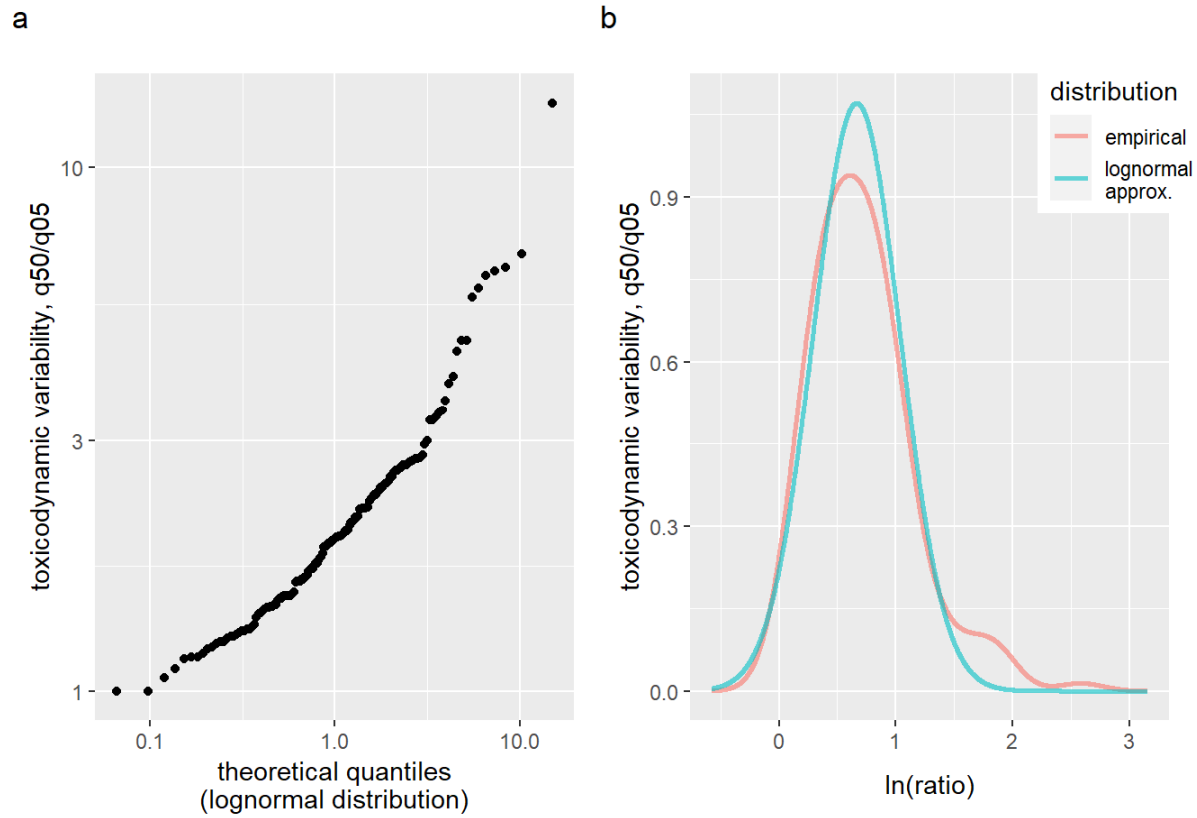


Figure 2-7 Lognormal approximation of the empirical distribution of corrected factors for interindividual variability with 5% incidence based on the data from Abdo et al. (2015). QQ-Plot of the empirical quantiles against the theoretical quantiles of a lognormal distribution (a). Density of the empirical distribution (b, red) and the corresponding parametrized lognormal distribution (b, blue).

The distribution of factors for 5% incidence is described by a lognormal distribution While the outer quantiles deviate slightly, the most important quantiles are described reasonably well by the parameters **$\mu = 0.66$ and $\sigma = 0.37$** (**Figure 2-7**).

2.5.2.2 Distribution corresponding to 1% incidence

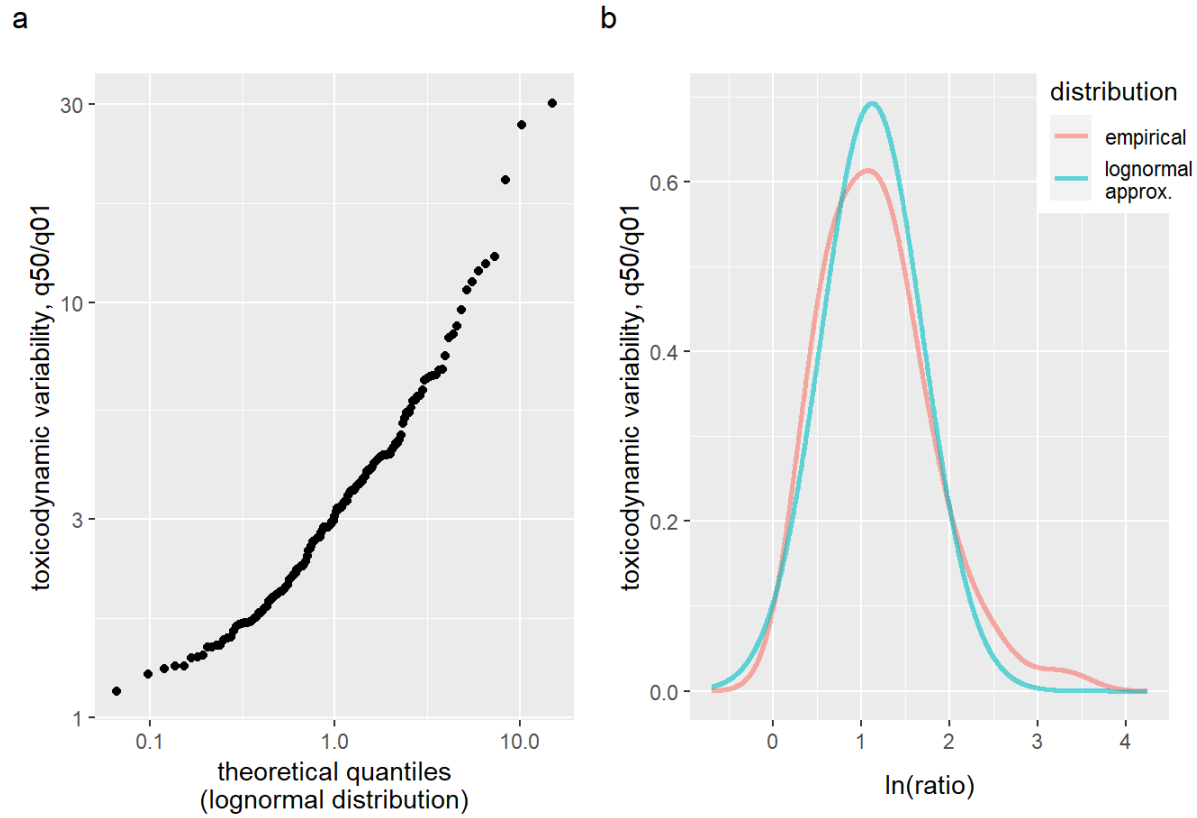


Figure 2-8 Lognormal approximation of the empirical distribution of corrected factors for interindividual variability with 1% incidence based on the data from Abdo et al. (2015). QQ-Plot of the empirical quantiles against the theoretical quantiles of a lognormal distribution (a). Density of the empirical distribution (b, red) and the corresponding parametrized lognormal distribution (b, blue).

The distribution of factors for 1% incidence can be described by a lognormal distribution. While the outer quantiles deviate slightly, the most important quantiles fit well to a lognormal distribution with the parameters **$\mu = 1.11$ and $\sigma = 0.58$** (**Figure 2-8**).

2.5.3 Combined intraspecies extrapolation

The toxicokinetic and toxicodynamic contribution to intraspecies variability is often combined to a single factor. For explorative reasons and to compare with default values for the combined factor (see section 3.3), the distributions for toxicokinetic and toxicodynamic variability corresponding to 5% and 1% incidence, respectively, are combined by Monte Carlo simulation. Based on the above, this is a multiplicative combination of the two distributions:

$$\text{Ratios}_{intra-combined,\,\text{i}} = \text{Ratios}_{intra-TK,\,\text{i}} * \text{Ratios}_{intra-TD,\,\text{i}} \qquad \text{(Equation 2)}$$

This ratio has the density

$$\text{pdf}(\text{Ratios}_{intra-combined,\,\text{i}}) = 10^{\left(\text{lognorm}_{\log_{10}\text{GSD}_{\text{TK}}}(\mu,\sigma)*z_{1-i}\right)} * \text{lognorm}_{\text{ratios}_{\text{TD}},\,\text{i}}(\mu,\sigma) \qquad \text{(Equation 3)}$$

Where i is the incidence and $z_{1\text{-}i}$ is the z-Score of the normal distribution corresponding to the fraction of the population to be covered. $\text{lognorm}_{\log10\,\text{GSD(TK)}}$ and $\text{lognorm}_{\text{ratios(TD)}}$ are the lognormal distributions characterized by μ and σ according to sections 2.5.1, 2.5.2.1 and 2.5.2.2.

The resulting simulated distribution of this combination ($1x10^7$ samples) is shown in **Figure 2-9** and **Table 2-7**. As expected, also with the combined distributions, the one covering 99% of the population requires larger factors than the one for 95% of the population. The parameters from **Table 2-7** are shown as the natural logarithms in **Figure 2-9**. For example, the median for the 95% distribution of 3.56 is located at 1.27 (ln 3.56 = 1.27) in the figure.

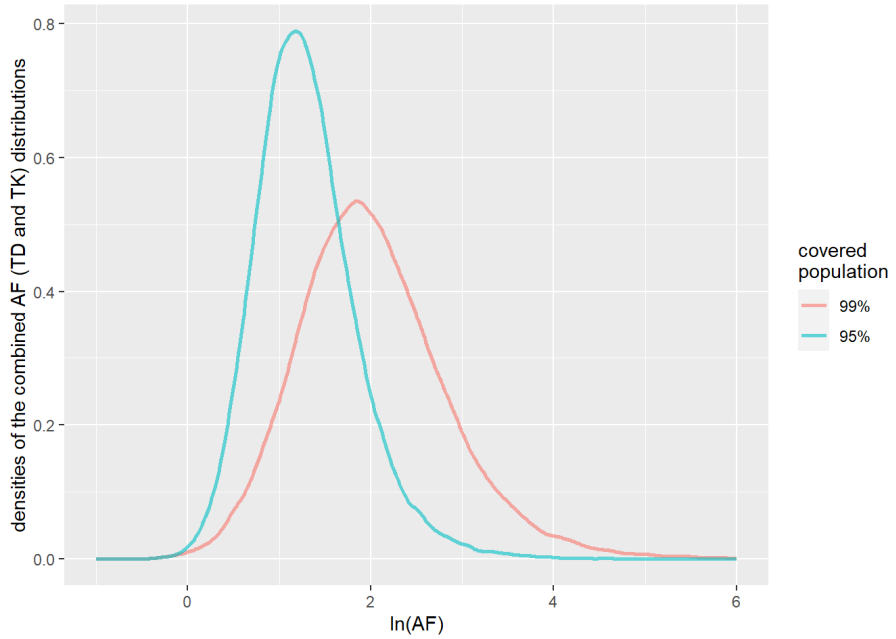


Figure 2-9 Distribution of the combined distribution for intraspecies variability, derived from Monte Carlo simulation of the multiplicative combination of distributions for toxicokinetic and toxicodynamic variability. Densities of the distributions corresponding to 1% incidence (red) and 5% incidence (blue) are plotted after logarithmic transformation

Table 2-7 Parameters of the combined distribution (TK and TD) for intraspecies variability for 1% and 5% incidence. Analysed by MC simulation of a multiplicative combination of distributions for TK and TD differences ($1x10^7$ samples each)

Incidence	GM	GSD	5%	Median	75%	95%
1 %	7.83	2.35	2.29	7.25	12.53	34.26
5 %	3.77	1.79	1.66	3.56	5.15	10.37

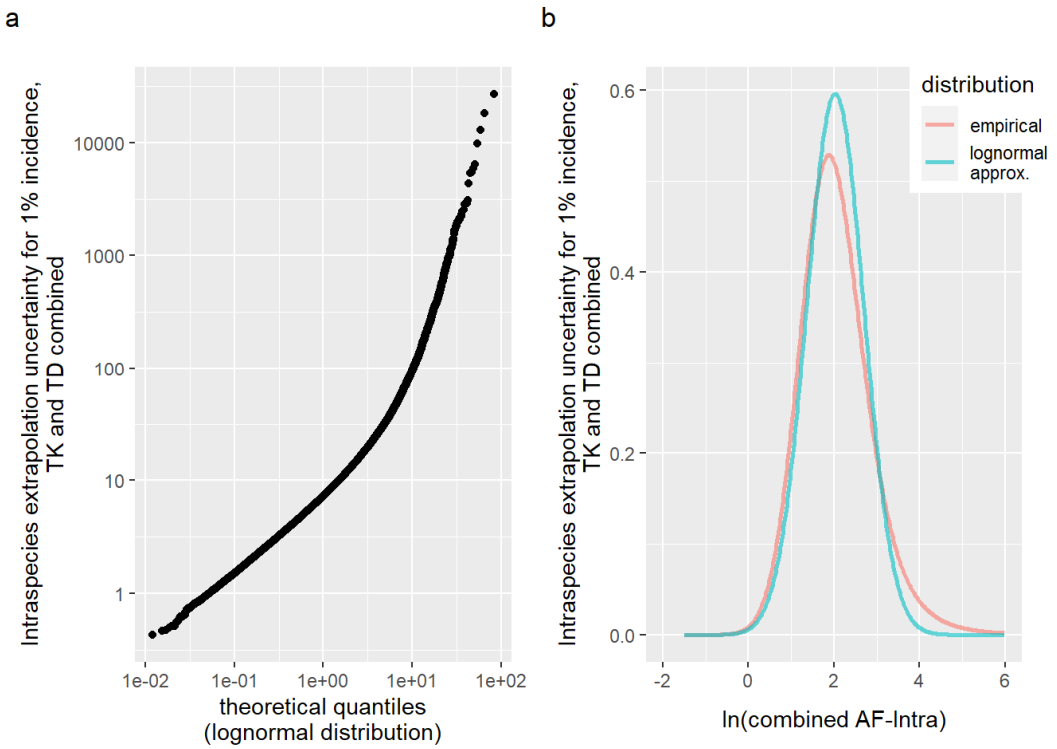


Figure 2-10 Lognormal approximation of the combined distribution of AF for intraspecies variability for a 1% incidence. QQ-Plot of the empirical quantiles against the theoretical quantiles of a lognormal distribution (a). Density of the empirical distribution (b, red) and the corresponding parametrized lognormal distribution (b, blue). The Monte Carlo simulation was performed with $1x10^7$ samples, but for computational reasons, the QQ-Plot is based on a subset of $1x10^5$ samples

Considering the nature of the combined distributions, it is not surprising that the resulting distribution is not well described by a lognormal distribution (exemplary shown in **Figure 2-10** for 1% incidence). However, analogous to the explanations given in

section 2.5.1 it is not necessary to parametrize the transformed distribution itself for the purpose of the analyses in this report.

2.6 Discussion

For all extrapolation steps and their respective empirical distributions obtained from the previous analyses parametric distributions were sought. QQ plots revealed that for all extrapolation steps except one, lognormal distributions were suitable. For intraspecies variability due to toxicokinetic differences the empirical distribution of log GSD values also resulted in a lognormal distribution. However, the final distribution is obtained according to the following equation

$$\mathrm{Ratio}_{\mathrm{Intra,\,i}} = 10^{\log_{10}\mathrm{GSD}*z_{1-i}}$$

This means that the final distribution is the basis 10 raised to the power of the lognormal distribution of log GSD (times z) and is not lognormal anymore. However, although not strictly lognormal the distribution is skewed to the right and of similar shape to a lognormal function (see Figure 2-10).

All distributions can easily be used in a probabilistic approach with tools such as EFSA's MonteCarlo tool.

In the following table the distributions obtained are compared with the distributions proposed in the WHO/IPCS report on uncertainty in hazard assessment (WHO 2014) and used in the APROBA (approximated probabilistic model, see also in the separate report on "Probabilistic hazard assessment") and the distributions developed in a previous project on a probabilistic model for hazard characterisation of chemicals at the workplace (Schneider et al. 2004; Schneider et al. 2006)).

Table 2-8 Comparison between parametric distributions as proposed by WHO (2014), Schneider and colleagues (Schneider et al. 2004; 2006) and this report

Extrapolation step	Percentiles	WHO/IPCS	Schneider et al.	This report
Time extrapolation: subacute – chronic	5th percentile	0.63		0.66
	50th percentile		4.14	3.71
	95th percentile	40	13.31	20.85
Time extrapolation: subchronic – chronic	5th percentile	0.5		0.55
	50th percentile		4.39	2.83
	95th percentile	8	11.74	14.49
Interspecies extrapolation	5th percentile	0.33		0.30
	50th percentile		0.97	1.02
	95th percentile	3	6.67	3.49
Intraspecies extrapolation: 5% incidence level	5th percentile	1.77		1.66
	50th percentile		4.82	3.56
	95th percentile	14.0	43.78	10.37
Intraspecies extrapolation: 1% incidence level	5th percentile	2.24		2.29
	50th percentile		9.96	7.25
	95th percentile	41.9	193.44	34.26

The comparison shows a high degree of concordance of the distributions proposed in this report with those in WHO (2014). Apart from our distribution for time extrapolation subchronic to chronic (which has a higher 95th percentile), all distributions in our report are similar (interspecies extrapolation) or slightly narrower (all others) than those in the WHO report. Different databases were used by WHO and in our study to derive the distributions for all extrapolation steps. This corroborates that these distributions can be assumed to represent the uncertainty and variability as observable in toxicological studies and human toxicokinetic studies.

The narrower distributions for intraspecies extrapolation (indicating less variability) are likely attributable to our use of the data by (Abdo et al. 2015) to derive a distribution for toxicodynamic variability (instead of the Hattis data used in the WHO report), in addition to a slightly narrower distribution for toxicokinetic differences. In fact, very little information on special risk groups such as children, elderly or sick people is included in the WHO distribution and the inclusion/non-inclusion of these groups is not considered to be a major cause of the difference in these distributions.

The distributions used to establish a probabilistic model for workplace chemicals in a previous research project for BAuA (Schneider et al. 2004; Schneider et al. 2006) provide higher medians but lower 95 percentile values for time extrapolation. The differences are likely caused by an improved database from evaluating a large set of NTP studies in this report. The other distributions are broader than those used here, most strikingly in the case of intraspecies extrapolation. Schneider et al. (2006) used the Hattis database, including the datasets with high variability from observations of immunological parameters, which were not considered reliable enough in this project and replaced by the new data provided by Abdo et al. (2015).

3 Comparison of distributions with currently used default values

3.1 Time extrapolation

An overview of assessment factors (AF) proposed in guidances for relevant regulatory frameworks is given in **Table 3-1**. For more details on factors used in various regulatory schemes we refer to the report “Comparison of methods for deriving OELs”.

3.1.1 Subacute – chronic

For the time extrapolation subacute – chronic all methodologies recommend an AF of 6 for systemic and local effects, except the ECETOC guidance, which recommends an AF of 1 for local effects (**Table 3-1**). As discussed in our report on “Time extrapolation”, our evaluation did not indicate a relevant difference between the distributions for local and for systemic effects. Therefore, we only derived parameters for a single distribution.

According to this distribution, an AF of 6 corresponds to the 67.7th percentile (see **Figure 3-1** and **Figure 3-2**) and the AF of 1 corresponds to 10.6% (Figure 3-2), which means that the commonly used factor of 6 achieves a rather high coverage of 68%, whereas ECETOC’s factor of 1 for local effects is expected to cover only 11% of cases. Therefore, when deriving OELs for new substances for which chronic studies are missing, time extrapolation with an AF of 6 would be conservative enough for 68% of newly assessed substances (with the implicit assumption that the new substances show the same behaviour as the substances evaluated in the NTP dataset from which the distribution was obtained).

For interpretation of these graphical presentations:

The obtained distribution is presented both as probability density function (PDF, blue, with left axis) and as cumulative distribution function (CDF, red, with right axis)

Vertical lines represent certain assessment factors used. The more to the right the vertical line is located, the higher the coverage achieved (probability that in a specific assessment the protection goal is achieved). The CDF directly allows to read out the probability achieved by a certain factor (intercept point of vertical line with CDF, read out on right axis).

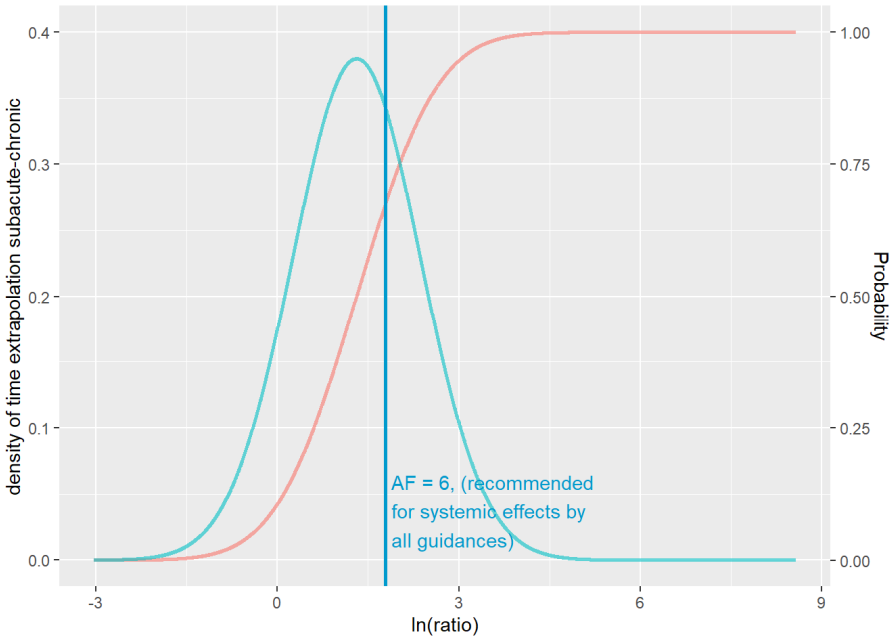


Figure 3-1 Coverage of the uncertainty for subacute/chronic extrapolation by commonly recommended assessment factors for **systemic** effects (all guidances recommend a factor of 6). The density of the distribution (blue, left axis) and the cumulative distribution (probability) (red, right axis) is shown

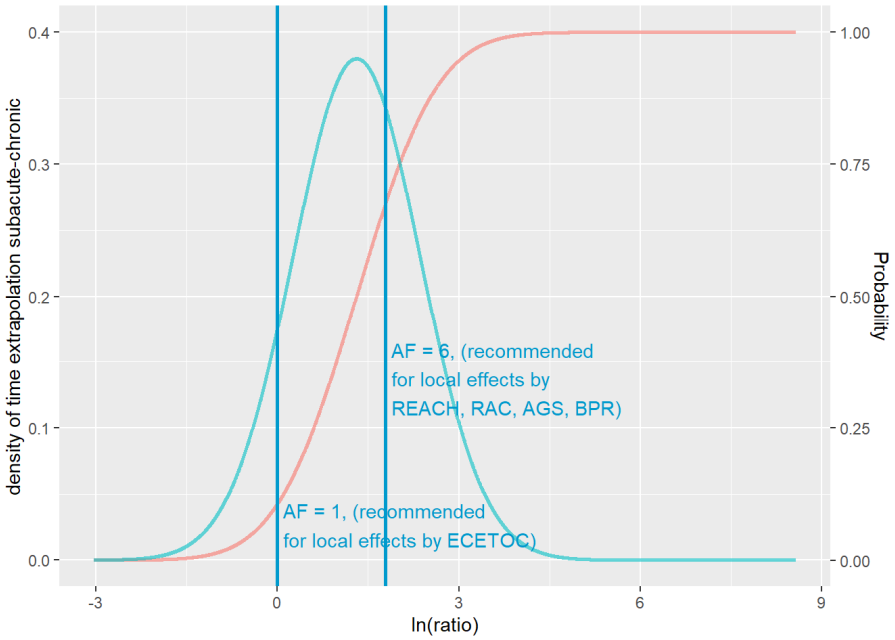


Figure 3-2 Coverage of the uncertainty for subacute/chronic extrapolation by commonly recommended assessment factors for **local** effects (usually 6, ECETOC recommends 1). The density of the distribution (blue, left axis) and the cumulative distribution (probability) (red, right axis) is shown

Table 3-1 Recommended factors for subacute/chronic time extrapolation according to relevant regulatory frameworks and their coverage according to the derived uncertainty distributions in this report

		REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**EU Plant Protection Products Directive**	**EU Biocidal Products Regulation**
systemic effects	Recommended factor	6	6	No default	6	No default	6	No default	6
	coverage by factor	67.7%	67.7%	-	67.7%	-	67.7%	-	67.7%
local effects	Recommended factor	6	No default	No explicit default	6	6 (sensory irritation)	1	No default	6 (reference to REACH)
	coverage by factor	67.7%	-		67.7%	67.7%	10.6%	-	67.7%

3.1.2 Subchronic – chronic

For the time extrapolation subchronic – chronic all methodologies recommend an AF of 2 for systemic and local effects, except the ECETOC guidance, which recommends an AF of 1 for local effects (**Table 3-2**). As discussed in the report on "Time extrapolation", our evaluation did not indicate a relevant difference between the distributions for local and for systemic effects. Therefore, we only derived parameters for a single distribution.

According to this distribution, an AF of 2 corresponds to a coverage of 36.3% (**Figure 3-3** and **Figure 3-4**) and the AF of 1 corresponds to 14.6% (**Figure 3-3**).

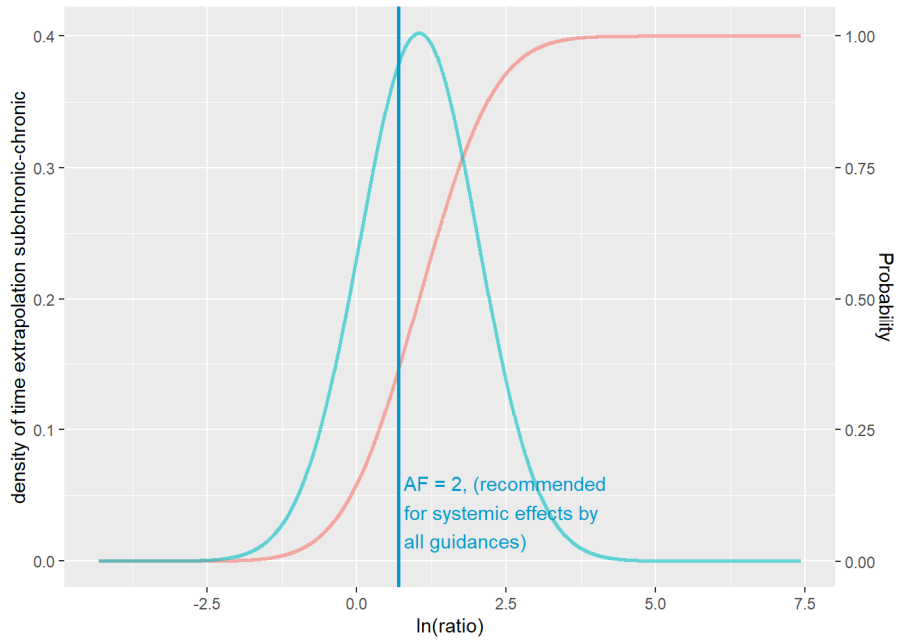


Figure 3-3 Coverage of the uncertainty for subchronic/chronic extrapolation by commonly recommended assessment factors for **systemic** effects (all guidances recommend a factor of 2). The density of the distribution (blue, left axis) and the cumulative distribution (probability) (red, right axis) is shown

The commonly applied factor of 2 for subchronic to chronic extrapolation yields a lower probability of 36% than the factor of 6 used for subacute to chronic extrapolation. A factor of approx. 3 would be required to achieve a similar probability as with the subacute – chronic factor. Again, ECETOC's factor of 1 for local effects leads to a low probability (15%).

Table 3-2 Recommended factors for subchronic/chronic time extrapolation according to relevant regulatory frameworks and their coverage according to the derived uncertainty distributions in this report.

		REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**EU Plant Protection Products Directive**	**EU Biocidal Products Regulation**
systemic effects	Recommended factor	2	2	No default	2	No default	2	2	2
	coverage by factor	36.3%	36.3%	-	36.3%	-	36.3%	36.3%	36.3%
local effects	Recommended factor	2	No default	No explicit default	2	2 (sensory irritation)	1	No default	2 (reference to REACH)
	coverage by factor	36.3%	-	-	36.3%	36.3%	14.6%	-	36.3%

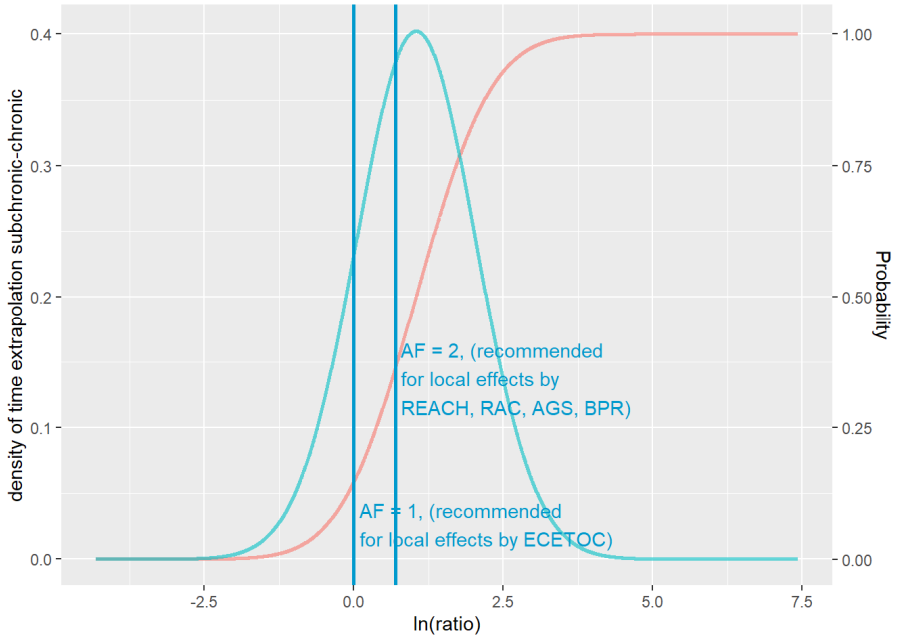


Figure 3-4 Coverage of the uncertainty for subchronic/chronic extrapolation by commonly recommended assessment factors for **local** effects (usually 6, ECETOC recommends 1). The density of the distribution (blue, left axis) and the cumulative distribution (probability) (red, right axis) is shown

3.2 Interspecies extrapolation

Recommendations for default AF to account for interspecies variability are relatively diverse across the methodologies. The defaults range from 1 (ECETOC) over 2.5 (REACH, RAC) for systemic effects (in addition to applying allometric scaling factors), and tend to be the same for local effects, if a default is provided (**Table 3-3**). The default value for the German OEL framework by AGS is a special case (combined factor for interspecies and intraspecies variability) and is addressed separately in section 3.4. PPP and BPR are special in this regard, as in these methodologies no allometric scaling is used. For a rat study with a caloric demand scaling factor of 4, the factor of 10 used in PPP and BPR would be equivalent to a factor of 2.5 for remaining differences and would therefore be the same as in the REACH guidance.

According to our evaluation, these default AF (after application of allometric scaling) correspond to coverage of 48.6% (AF=1) and 88.4% (AF=2.5) (**Figure 3-5** and **Figure 3-6**). It should be noted that the distribution, which the defaults are compared to, is based on the corrected σ, thus corresponds to the uncertainty for interspecies extrapolation without the variance introduced by uncertainties inherent to the NOAEL or BMD. For comparison, the coverages of the AF 1, and 2.5 of the uncorrected distribution for interspecies differences obtained from the NTP data evaluation are 49.1% and 77.8% (no figure shown).

Overall, the commonly used factor of 2.5 (applied in combination with allometric scaling) is covering a large part of cases (coverage 88%), whereas the factor of 1 accounts for approx. half of the cases, which is the expected result, because after

allometric scaling on average experimental animals and humans are expected to be equally susceptible. The interspecies factor of 10 applied in the frame of PPP and BPR would lead to the same protection goal as the factor of 2.5 (plus allometric scaling) in the case of a rat study, but would be more protective in case of larger experimental animals (e.g. dog) and less protective with smaller animals (e.g. mice).

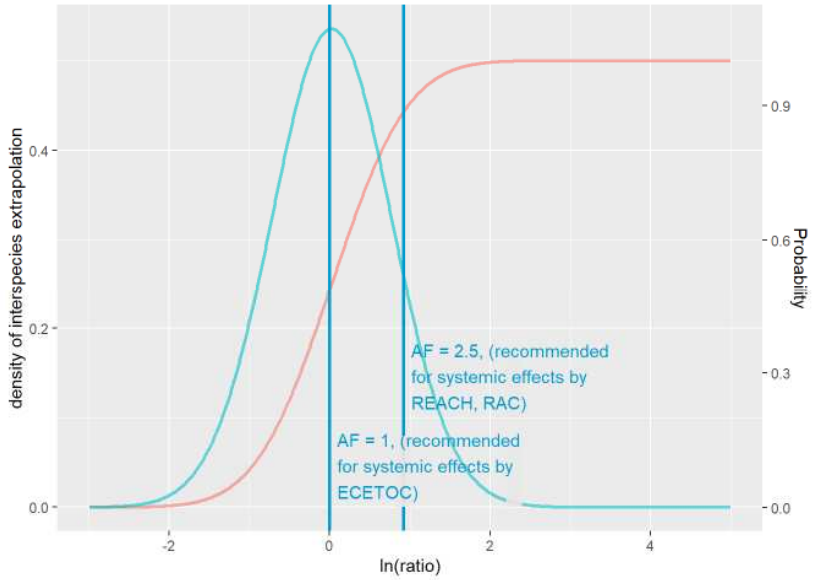


Figure 3-5 Coverage of the uncertainty for interspecies extrapolation by commonly recommended assessment factors for **systemic** effects (ECETOC: 1, REACH, RAC: 2.5). The density of the distribution (blue, left axis) and the cumulative distribution (probability) (red, right axis) is shown

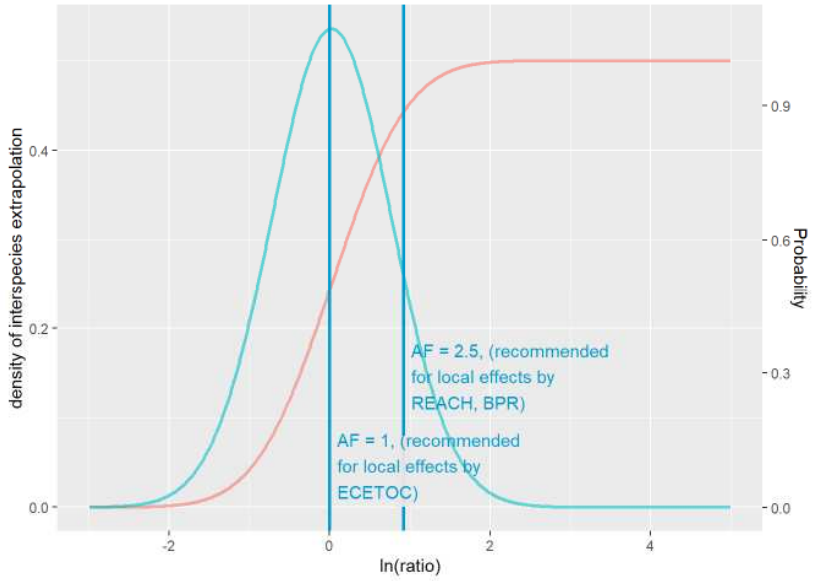


Figure 3-6 Coverage of the uncertainty for interspecies extrapolation by commonly recommended assessment factors for **local** effects (ECETOC: 1, REACH, BPR: 2.5). The density of the distribution (blue, left axis) and the cumulative distribution (probability) (red, right axis) is shown

Table 3-3 Recommended factors for interspecies extrapolation according to relevant regulatory frameworks and their coverage of the population according to the derived uncertainty distributions in this report.

		REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**EU Plant Protection Products Directive**	**EU Biocidal Products Regulation**
systemic effects	Recommended factor	2.5	2.5	No default provided	Inter + Intra = 5	No default provided	1	10 (without scaling)	10 (without scaling)
	coverage by factor	88.4%	88.4%	-	See section 3.4	-	48.6%	-	-
local effects	Recommended factor	2.5	No default provided	Specific for sensor. irritation	Inter + Intra = 5	3 (for sensory irritation only)	1	-	2.5
	coverage by factor	88.4%	-	-	See section 3.4	92.5%	48.6%	-	88.4%

3.3 Intraspecies extrapolation

The evaluation of coverage is based on the combined assessment factor for toxicodynamic and toxicokinetic differences, as all default values given in OEL frameworks refer to the combined factor.

The defaults for workers range from 2 (SCOEL for local irritation), 3 (ECETOC) over 5 (REACH, RAC) up to 10 (PPP, BPR). Generally, no differences exist between local and systemic effects, yet the SCOEL methodology provides an unbounded default for systemic effects, with a lower bound below the default for sensory irritation, and the RAC and PPP methodology provide defaults only for systemic effects (**Table 3-4**). For the special case of the AGS framework (combined factor for interspecies and intraspecies variability) see section 3.4. The coverage of the distribution for intraspecies extrapolation by these default AF is analysed for the two scenarios 1% and 5% incidence in the population. Assessment factors are compared to the distribution for combined TK and TD intraspecies variability, as developed in section 2.5.3.

In the case of a 1% incidence level, the default AF of 2, 3, 5 and 10 correspond to a coverage of 3.2%, 10.8%, 30.7% and 65.8% (Table 3-4 and upper panels of Figure 3-7 and Figure 3-8). For the distribution corresponding to a 5% incidence, these default factors cover 11.1%, 36.6%, 73.4% and 94.6% (Table 3-4 and lower panels of Figure 3-7 and Figure 3-8).

This leads to the interpretation that for the objective to cover 95% of exposed workers (incidence 5%) the factor of 5 proposed in the REACH guidance for workers (for both local and systemic effects) provides for a coverage of about 73% of cases, whereas the factor of 10 proposed in the BPR guidance achieves a high probability of 95%. If, however, the objective is to protect 99% of workers, then the factor of 5 is sufficient only in 31% of cases and even the factor of 10 achieves only a probability of 66%.

The factor of 3 proposed by ECECTOC leads to probabilities of 11 and 37% only, for 1% and 5% incidence, respectively. The SCOEL factor of 2 for sensory irritation leads to even lower probabilities. However, it must be noted that our distribution does not specifically consider sensory irritation as an endpoint. A sound interpretation of the factor proposed by SCOEL is therefore not possible.

In conclusion, except the factor of 10 in the BPR guidance, currently used assessment factors for intraspecies variability provide for a rather low probability (i.e. low percentage of cases covered).

Table 3-4 Recommended factors for intraspecies extrapolation according to relevant regulatory frameworks and their coverage of the population according to the derived uncertainty distributions in this report.

		REACH Regulation	**RAC OEL metho-dology**	**SCOEL**	**AGS – German OELs**	**DFG MAK**	**ECETOC**	**EU Plant Protection Products Directive**	**EU Biocidal Products Regulation**
systemic effects	Recommended factor	5	5	>= 1	Inter + Intra = 5	No default provided	3	10	10
	coverage by factor for 1% incidence	30.7%	30.7%	-	-	-	10.8%	65.8%	65.8%
	coverage by factor for 5% incidence	73.4%	73.4%	-	-	-	36.6%	94.6%	94.6%
local effects	Recommended factor	Worker: 5	No default provided	2 for sensor. irritation, no default for other local effects	Inter + Intra = 5	No default provided	Worker: 3	-	10 (for professionals and non-professionals)
	coverage by factor for 1% incidence	30.7%	-	3.2%	-	-	10.8%	-	65.8%
	coverage by factor for 5% incidence	73.4%	-	11.1%	-	-	36.6%	-	94.6%

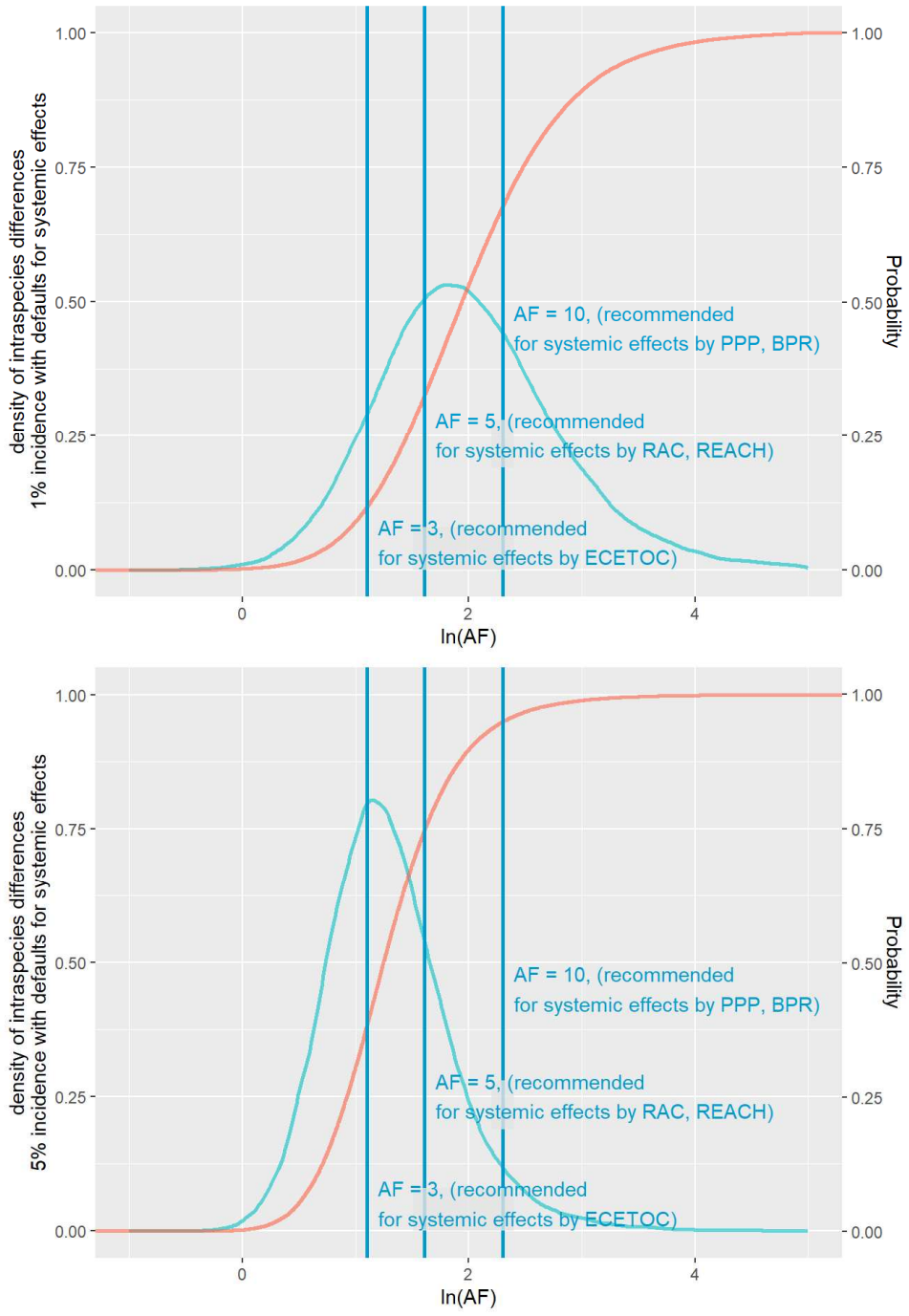


Figure 3-7 Coverage of the uncertainty for intraspecies extrapolation by commonly recommended assessment factors for **systemic** effects (ECETOC: 3, REACH, RAC: 5, PPP, BPR: 10). The density of the distribution (blue, left axis) and the cumulative distribution (probability) (red, right axis) is shown for an incidence level of 1% (upper plot) and 5% (lower plot)

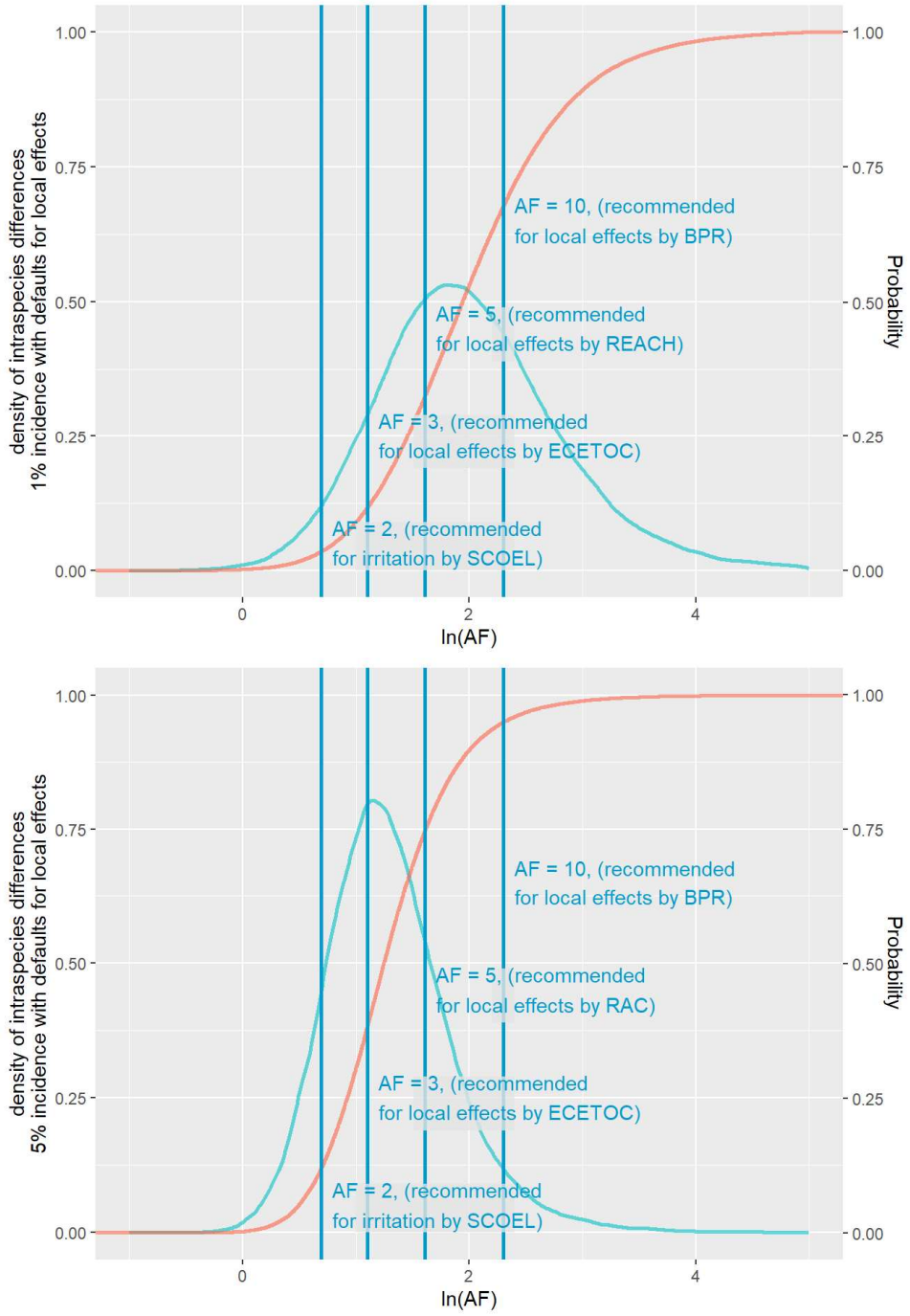


Figure 3-8 Coverage of the uncertainty for intraspecies extrapolation by commonly recommended assessment factors for **local** effects (SCOEL: 2 (sensory irritation only), ECETOC: 3, REACH: 5, BPR: 10). The density of the distribution (blue, left axis) and the cumulative distribution (probability) (red, right axis) is shown for an incidence level of 1% (upper plot) and 5% (lower plot)

3.4 Combined factor for inter- and intraspecies variability used for deriving OELs in Germany

The German OEL framework by AGS recommends a default AF of 5 to cover the combined uncertainty due to inter- and intraspecies differences. In order to evaluate the corresponding coverage provided by this factor, the distributions derived in this report to cover inter- and intraspecies differences are combined by multiplication using a MC simulation with $1x10^7$ samples. For the intraspecies uncertainty, the distributions corresponding to 1% and 5% incidence are used.

In order to illustrate the consequences of such a combination of two distributions, the involved distributions for the MC simulation for 1% are shown in detail in **Figure 3-9**. In this figure the distribution for interspecies (top) is aligned to the intraspecies distribution (middle) in such way that the expected values (μ) coincide on the x-axis in order to aid understanding of the procedure. Multiplication of the two distributions on the normal scale corresponds to an addition of the two distributions on the depicted ln-scale. This results in a higher dispersion of the distribution for the combined inter- and intraspecies uncertainty (bottom plot). Because of the increased dispersion, percentiles below the median are at lower ratios than for the individual distributions and correspondingly, percentiles above the median are at higher ratios than for the individual distributions.

Thus, combining the two distributions does not change the central position of the intraspecies distribution, but adding the uncertainty of the interspecies extrapolation broadens its width.

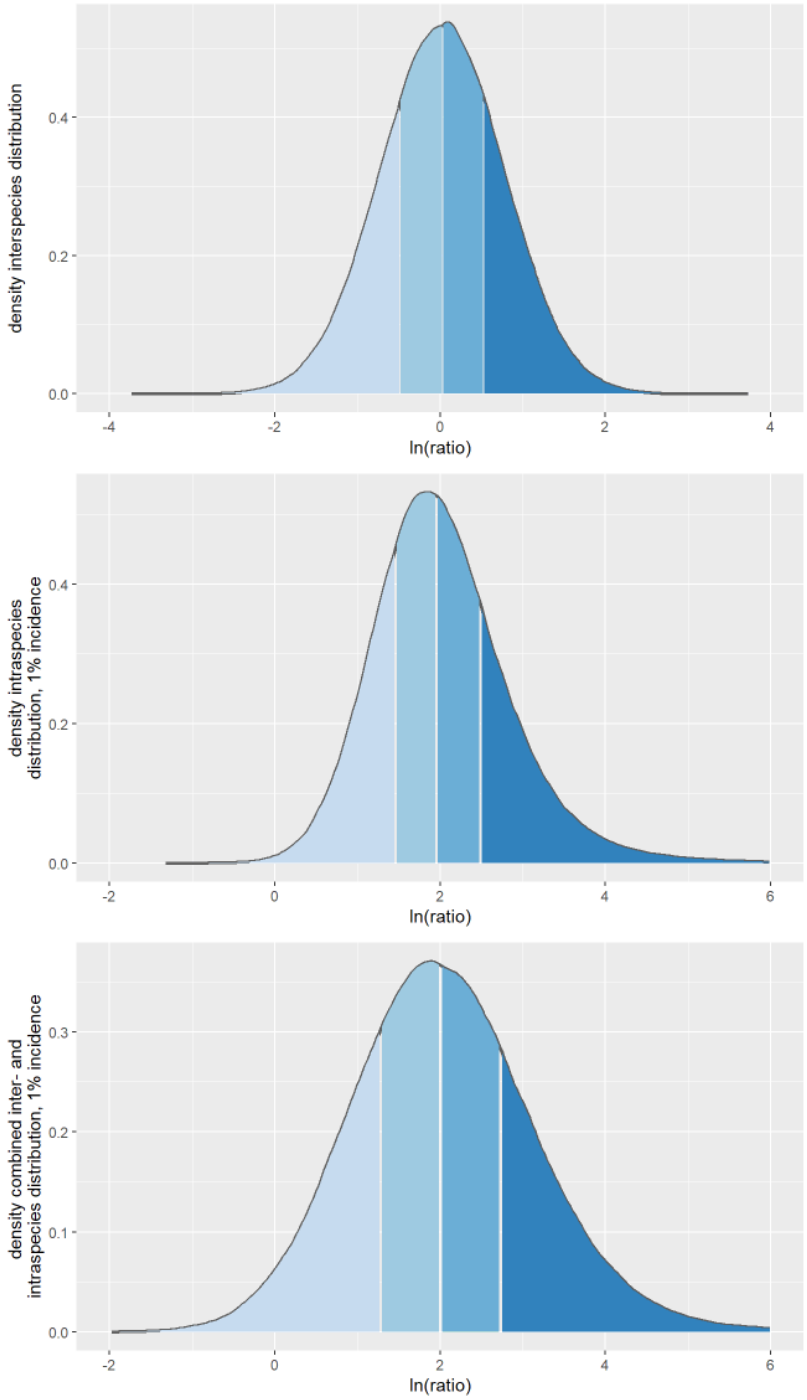


Figure 3-9 Monte Carlo simulation of the multiplication of the interspecies (top) with intraspecies (middle) distribution for 1% incidence and the resulting distribution for the combination of inter- and intraspecies differences (bottom). Please note that the x-axis of the interspecies distribution is aligned to match the central moment of the intraspecies distribution, but all three x-axes span 8 ln units in order to be comparable. The shades of blue correspond to 25%, 50%, 75% and 100% of the population.

Parameters of the distribution and the coverage of the default AF of 5 for both an incidence of 1% and 5% are shown in **Figure 3-10** and **Table 3-5**.

Table 3-5 Parameters of the combined distribution for intra- and interspecies variability analysed by MC simulation of a multiplicative combination of AF for inter- and intraspecies (1% and 5% incidence) differences ($1x10^7$ samples each).

Incidence	**GM**	**GSD**	**5%**	**Median**	**75%**	**95%**
1%	8.03	3.11	1.38	7.66	16.22	54.09
5%	3.87	2.57	0.86	3.77	7.09	18.77

According to our evaluation, the AF of 5 covers 34.5% of the distribution in case an incidence of 1% is targeted and provides a 61.9% coverage for a 5% incidence level. Hence, the interpretation is: The combined factor of 5 is sufficient to provide protection

- in 62% of cases (evaluations, substances), if the aim is to include 95% of workers (5% incidence), and
- in 35% of cases, if the evaluation aims at protecting 99% of workers (1% incidence).

The relatively low probability is mainly caused by the distribution for intraspecies variability. In section 3.3 we already described that for intraspecies variability a factor of 5 is required for a probability of approx. 73% at the 5% incidence level and a factor of 10 is necessary for 95% probability. This means that the probability achieved by a factor of 5 for intraspecies extrapolation alone is only slightly higher (73%) compared to the probability achieved, when the factor of 5 is used as a combined factor for inter- and intraspecies extrapolation (62%). A factor of 7 would be required to achieve a probability of 75% at the 5% incidence level for the combined factor and even higher factors at the 1% incidence level (Table 3-5).

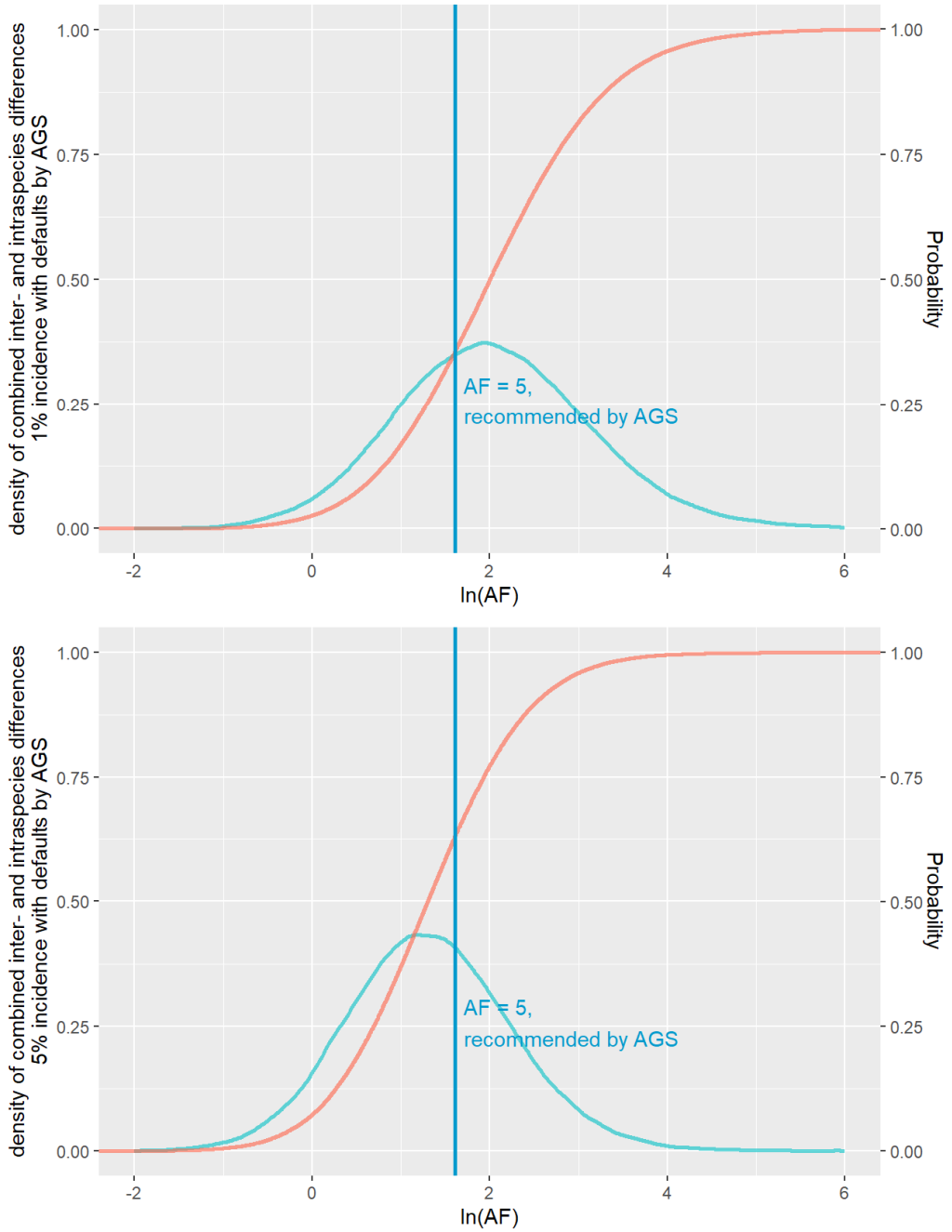


Figure 3-10 Coverage of the default AF of 5 regarding the combined variability of factors for intra- and interspecies differences, derived by Monte Carlo simulation of the multiplicative combination of AF for inter- and intraspecies differences. The density (blue) and cumulative probability (red) of the distribution with an incidence of 1% (upper plot) and 5% (lower plot) for the intraspecies variability is plotted after logarithmic transformation.

3.5 Distributions for all extrapolation steps and comparison with existing methodologies

In the following, the coverage of the proposed AF in relevant OEL frameworks according to the combined uncertainty distributions are exemplary evaluated for the following combination of extrapolation steps

- Time extrapolation: subacute study, i.e. subacute to chronic extrapolation
- Interspecies extrapolation: Species is rat, i.e. an allometric scaling factor of 4 is applied as single numerical value, together with the distribution for remaining species differences
- Intraspecies extrapolation: calculations are done at the 5% incidence (coverage of 95% of workers) and the 1% incidence level (coverage of 99% of workers)

The respective distributions are combined by Monte Carlo simulation according to Equation 4 and 5 for systemic and local effects, respectively.

$$combined\ distribution_{\text{systemic effects}} = distribution_{\text{time}} * allometric\ scaling * distribution_{interspecies} * distribution_{\text{intraspecies}} \quad \text{(Equation 4)}$$

$$combined\ distribution_{\text{local effects}} = distribution_{\text{time}} * distribution_{interspecies} * distribution_{\text{intraspecies}} \quad \text{(Equation 5)}$$

(assuming an air concentration as POD with no need for allometric scaling)

The resulting simulated distribution (see Fig 3-11 for the 5% incidence level) is compared to the assessment factors used by the following frameworks for both systemic and local effects:

- RAC/REACH,
- AGS,
- EU BPR and
- ECETOC.

For the other frameworks default factors are not available for some of the individual extrapolation steps.

For an assessment, which starts with a POD from an subacute study, the total AF and the corresponding coverage according to the probabilistic modelling is given in **Table 3-6**. The probabilistic modelling is performed with intraspecies distributions corresponding to 1% and 5% incidence. Probabilities (coverage) are calculated separately for systemic effects observed in an oral rat study (including use of an allometric scaling factor 4), for systemic effects observed in a subacute inhalation study and for local effects in a subacute inhalation study.

Table 3-6 Probability achieved (covered fraction of the uncertainty distribution according to probabilistic modelling using the parameters presented in this report) by the default AF proposed in relevant OEL frameworks (POD from subacute study).

OEL framework	**type of effect, route of application**	**Proposed AF (time, inter, scaling, intra)**	**total AF proposed**	**incidence for Intraspecies factor**	**coverage by total AF**
REACH/RAC	systemic, oral	6, 2.5, 4, 5	300	1 %	**73.3 %**
				5 %	**88.0 %**
	systemic, inhalation	6, 2.5, -, 5	75	1 %	**73.3 %**
				5 %	**88.0 %**
	local, inhalation	6, 2.5, -, 5	75	1 %	**73.3 %**
				5 %	**88.0 %**
AGS	systemic, oral	6, 5, 4, -	120	1 %	**51.0 %**
				5 %	**70.3 %**
	systemic, inhalation	6, 5, -, -	30	1 %	**51.0 %**
				5 %	**70.3 %**
	local, inhalation	6, 5, -, -	30	1 %	**51.0 %**
				5 %	**70.3 %**
EU BPR	systemic, oral	6, 10, -, 10	600	1 %	**85.6 %**
				5 %	**95.0 %**
	systemic, inhalation*				
	local, inhalation	6, 2.5, -, 10	150	1 %	**85.6 %**
				5 %	**95.0 %**
ECETOC	systemic, oral	6, 1, 4, 3	72	1 %	**37.7 %**
				5 %	**56.8 %**
	systemic, inhalation	6, 1, -, 3	18	1 %	**37.7 %**
				5 %	**56.8 %**
	local, inhalation	1, 1, -, 3	3	1 %	**6.5 %**
				5 %	**13.3 %**

* inhalative PODs are converted to systemic dose descriptors, same AF as for systemic, oral apply

The following **Figure 3-11** visualises these differences in protection goals achieved by default factors used in the various OEL frameworks.

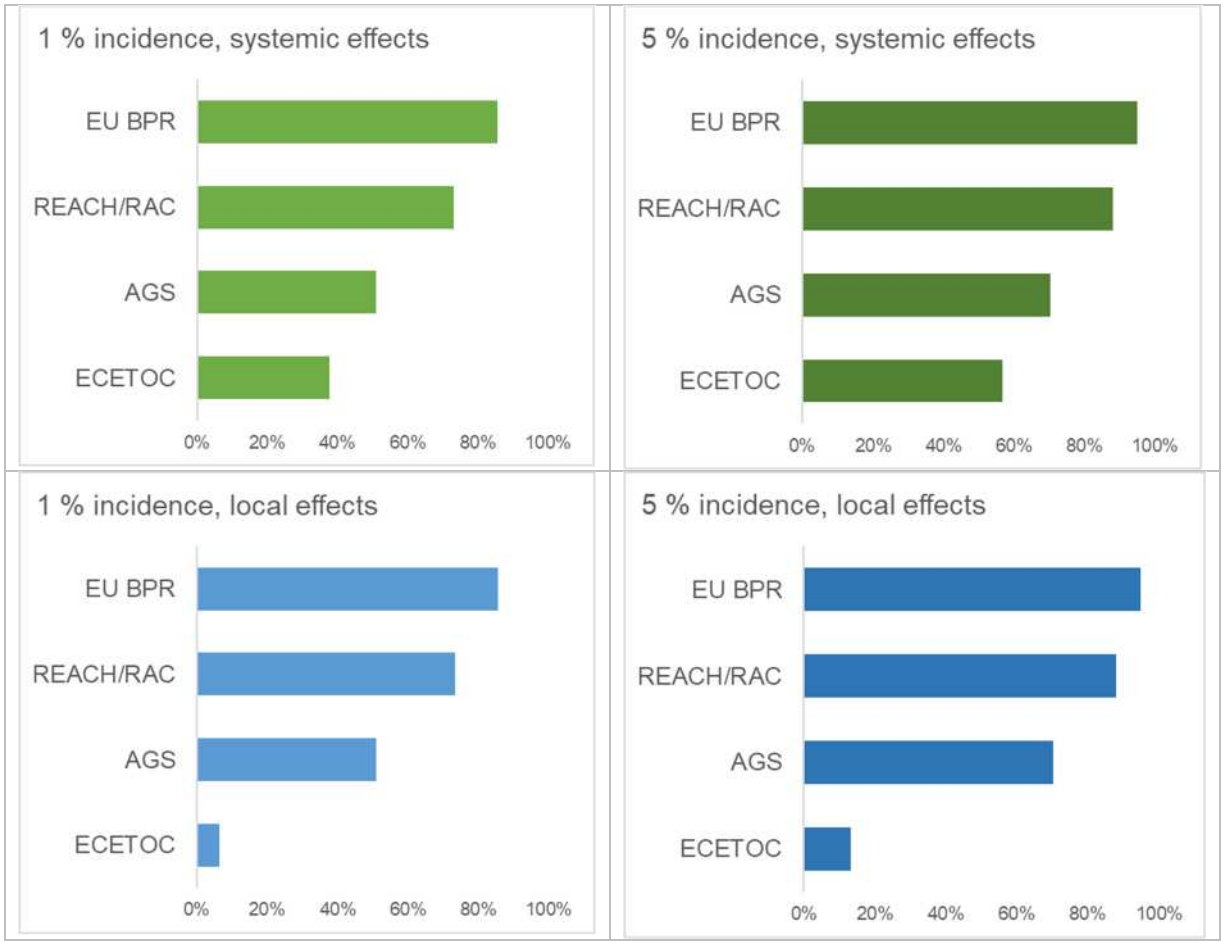


Figure 3-11 Probabilities (coverage) achieved by the default factors applied in different OEL frameworks (POD from subacute study)

In **Figure 3-12** the distributions obtained from probabilistically combining the distributions of individual extrapolation steps are presented graphically for the 5% incidence level exemplary, for systemic and local effects and compared to the combined default values of various OEL frameworks (vertical lines).

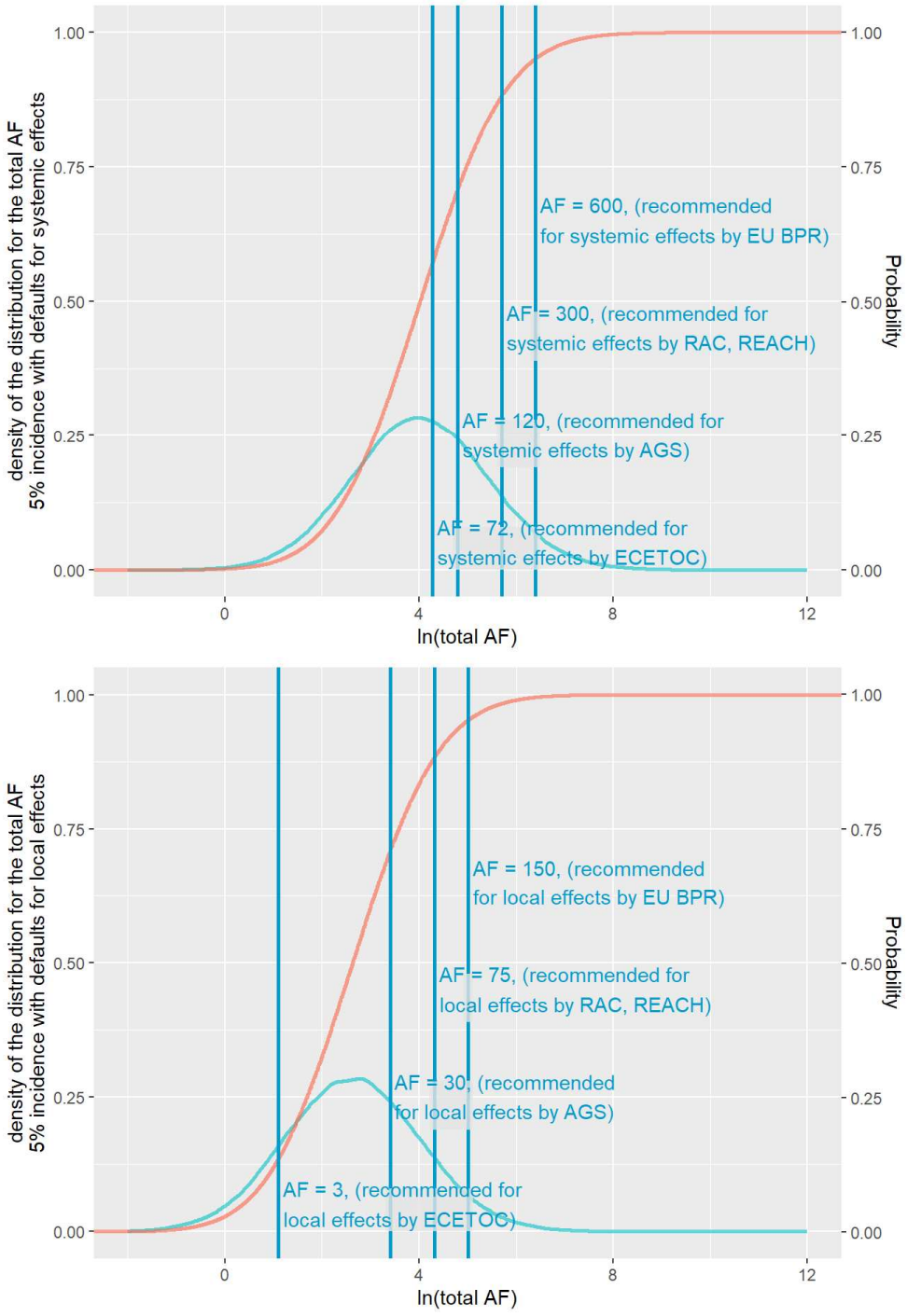


Figure 3-12 Coverage of the default AF for the scenario of a subacute rat study with systemic or local effects. Covered probabilities are derived by Monte Carlo simulation according to Equation 4 (systemic effects) and Equation 5 (local effects) using the intraspecies distribution corresponding to 5% incidence. The density (blue) and cumulative probability (red) of the distribution is plotted after logarithmic transformation of the combined default AF as recommended by the OEL frameworks.

The scenario used for calculating the probability values in **Table 3-6** assumes a subacute study in rats as the key study. For a case requiring subchronic to chronic extrapolation (POD derived from a subchronic rat study) or for assessments starting from a chronic study (no time extrapolation) the respective probabilities are given in **Table 3-7** and **Table 3-8**.

Table 3-7 Probability achieved (covered fraction of the uncertainty distribution according to probabilistic modelling using the parameters presented in this report) by the default AF proposed in relevant OEL frameworks (POD from subchronic study)

OEL framework	**type of effect, route of application**	**Proposed AF (time, inter, scaling, intra)**	**total AF proposed**	**incidence for Intraspecies factor**	**coverage by total AF**
REACH/RAC	systemic, oral	2, 2.5, 4, 5	100	1 %	**53.3 %**
				5 %	**73.0 %**
	systemic, inhalation	2, 2.5, -, 5	25	1 %	**53.3 %**
				5 %	**73.0 %**
	local, inhalation	2, 2.5, -, 5	25	1 %	**53.3 %**
				5 %	**73.0 %**
AGS	systemic, oral	2, 5, 4, -	40	1 %	**29.5 %**
				5 %	**47.7 %**
	systemic, inhalation	2, 5, -, -	10	1 %	**29.5 %**
				5 %	**47.7 %**
	local, inhalation	2, 5, -, -	10	1 %	**29.5 %**
				5 %	**47.7 %**
EU BPR	systemic, oral	2, 10, -, 10	200	1 %	**70.8 %**
				5 %	**86.7 %**
	systemic, inhalation*				
	local, inhalation	2, 2.5, -, 10	50	1 %	**70.8 %**
				5 %	**86.7 %**
ECETOC	systemic, oral	2, 1, 4, 3	24	1 %	**18.7 %**
				5 %	**33.2 %**
	systemic, inhalation	2, 1, -, 3	6	1 %	**18.7 %**
				5 %	**33.2 %**
	local, inhalation	1, 1, -, 3	3	1 %	**8.5 %**
				5 %	**17.1 %**

* inhalative PODs are converted to systemic dose descriptors, same AF as for systemic, oral apply

Table 3-8 Probability achieved (covered fraction of the uncertainty distribution according to probabilistic modelling using the parameters presented in this report) by the default AF proposed in relevant OEL frameworks (POD from chronic study).

OEL framework	type of effect, route of application	Proposed AF (time, inter, scaling, intra)	total AF proposed	incidence for Intraspecies factor	coverage by total AF
REACH/RAC	systemic, oral	-, 2.5, 4, 5	50	1 %	**67.2 %**
				5 %	**89.5 %**
	systemic, inhalation	-, 2.5, -, 5	12.5	1 %	**67.2 %**
				5 %	**89.5 %**
	local, inhalation	-, 2.5, -, 5	12.5	1 %	**67.2 %**
				5 %	**89.5 %**
AGS	systemic, oral	-, 5, 4, -	20	1 %	**34.6 %**
				5 %	**62.0 %**
	systemic, inhalation	-, 5, -, -	5	1 %	**34.6 %**
				5 %	**62.0 %**
	local, inhalation	-, 5, -, -	5	1 %	**34.6 %**
				5 %	**62.0 %**
EU BPR	systemic, oral	-, 10, -, 10	100	1 %	**85.2 %**
				5 %	**97.2 %**
	systemic, inhalation*				
	local, inhalation	-, 2.5, -, 10	25	1 %	**85.2 %**
				5 %	**97.2 %**
ECETOC	systemic, oral	-, 1, 4, 3	12	1 %	**18.8 %**
				5 %	**40.2 %**
	systemic, inhalation	-, 1, -, 3	3	1 %	**18.8 %**
				5 %	**40.2 %**
	local, inhalation	-, 1, -, 3	3	1 %	**18.8 %**
				5 %	**40.2 %**

* inhalative PODs are converted to systemic dose descriptors, AF as above for oral apply

The achieved probabilities are lower when starting from a subchronic study (compared to a subacute study), as the currently used default values correspond to a lower percentile of the respective distribution for subchronic to chronic extrapolation compared to subacute to chronic extrapolation (see section 2.3). An exception is the ECETOC recommendation for locally acting substances, where an AF of 1 is proposed for all time extrapolations. Here, a slightly higher probability is achieved for the "subchronic case", because the "error of not using a time extrapolation" is less severe for subchronic-chronic extrapolation. Overall, the probabilities obtained for ECETOC assessment factors are low in all three cases.

With a POD from a chronic study, similar probabilities are obtained as for a POD from a subacute study (except for ECETOC, local inhalation, for which the probabilities are higher for the chronic study duration).

3.6 Factors needed to achieve a certain protection goal

In order to determine the total assessment factor needed to achieve a predefined protection goal (e.g. asking which factor is needed to achieve a probability 95% to protect from an effect when departing from a subacute study and aiming for 1% incidence in the worker population) the quantiles corresponding to the desired protection goals have to be determined for the respective combined distribution. This distribution results from the combination of distributions for the individual extrapolation steps. In Table 3-9 the resulting factors are given for three different probabilities: 50%, as an example of a probability that we expect would be perceived as low, as well as 75% and 95%, which probably better reflect regulatory practice. These examples do not include a potential allometric correction (when starting from oral studies) or other modifications of the PoD.

Table 3-9 Total extrapolation factors needed to achieve a defined probability (50%, 75% and 95%) for the scenarios 1% or 5% incidence and departing from a subacute, subchronic or chronic study.

Duration of key study	Incidence	Probability	Total AF needed
subacute	1%	50%	28.9
		75 %	81.5
		95 %	389
	5%	50%	14.1
		75 %	36.6
		95 %	149
subchronic	1%	50%	22.1
		75 %	60.6
		95 %	280
	5%	50%	10.8
		75 %	27.2
		95 %	106
chronic	1%	50%	7.6
		75 %	16.2
		95 %	54.1
	5%	50%	3.8
		75 %	7.1
		95 %	18.8

3.7 Discussion

The coverage provided by the currently used assessment factors (i.e. the probability of being on the safe side in a specific case or substance evaluation) is different for the various extrapolation steps:

Time extrapolation

- The extrapolation factor subacute to chronic of 6 used by several organisations results in a probability of 68%
- However, the coverage is less (only 36%) for the assessment factor of 2 used often for subchronic to chronic extrapolation

Interspecies extrapolation:

- The interspecies extrapolation factor of 2.5 (for systemic effects applied in combination with allometric scaling to cover remaining uncertainty) provides for a (high) probability of 88%
- The interspecies factor of 10 applied in the PPP and BPR framework (without allometric scaling) is achieving the same level of coverage in case of rat studies; for smaller species the probability would be lower, for larger species higher

Intraspecies extrapolation:

- For intraspecies extrapolation the coverage depends on how the target population to be considered by the OEL is defined; for our comparisons we used two levels: distributions for including 95% (5% incidence level) or 99% (1% incidence level) of the exposed population
- A high coverage was only achieved **at the 5% incidence level** by the factor of 10 used in PPP and BPR (95%) and - with a somewhat lower probability of 73% - by the factor of 5 used for REACH DNELs and RAC OELs.
- **At the 1% incidence level** the probabilities achieved by these factors were only 66% and 31%; other factors (proposed by ECETOC and SCOEL) yielded even lower probabilities.

This is the extrapolation step with the largest differences in assessment factors in current methodologies, which might well be due to the lack of a sound empirical database. This project provides an improved database for describing intraspecies variability. The data provided in our evaluation should be used to specifically check the approach taken for covering interindividual variability.

The German AGS methodology uses a combined factor for interspecies and intraspecies extrapolation. Combining the distributions for interspecies and intraspecies extrapolation essentially led to a distribution which is similar to the intraspecies distribution, but with a larger spread (as the GM of the interspecies distribution is 1 no shift of the intraspecies distribution in any direction results). This results in a higher uncertainty from the combined steps. The factor of 5 used in the AGS methodology leads to a coverage of 62% and 35% of cases at the 5% and 1% incidence level, respectively.

Total Assessment factor

The methodologies differ in some of the default assessment factors. When all distributions are combined by Monte Carlo analysis and compared with the product of the assessment factors of each methodology, the following sequence (from higher to lower probability) results:

BPR > RAC/REACH > AGS > ECETOC.

Achieved probabilities range from 95% to 57% at the 5% incidence level and from 86% to 38% at the 1% incidence level. For ECETOC, in the case of local effects, low probabilities of 13 and 7% (for the 5% and 1% incidence level) result.

As explained in section 3.5 this comparison is possible only for those methodologies providing default values for all extrapolation steps. However, it can be approximated that PPP would range similarly to BPR. The German MAK Commission's[5] methodology is expected to lead to lower probabilities compared to AGS, as generally a lower inter-/intraspecies factor is used.

Assessment factors proposed by ECETOC always result in the lowest coverage. This is especially true for local effects, where ECETOC assumes that effects are exclusively concentration-dependent, which is not in agreement with our findings.

When deriving individual assessment factors from these distributions, a decision on the desired coverage needs to be taken. It needs to be borne in mind that a combination of median values from each distribution will result again in probabilities around 50%. Combination of higher percentiles from each distribution would result in a higher percentile (than the individual ones) of the combined distribution. For example, combining 75th percentile values resulted in values around the 88th percentile in some of our example calculations.

[5] Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area (MAK Commission) of the Deutsche Forschungsgemeinschaft DFG

4 Uncertainty associated with the POD

4.1 Distribution of the BMD

All points of departure (POD) have an inherent uncertainty, which comes from

- the uncertainty with which effects can be observed in the experimental model or human cohort and
- the preciseness with which the POD determined from the observed data reflects the point of the "true" dose-response function adhering to the predetermined effect level (benchmark response).

One (of several) advantages of the benchmark method (see separate report "Benchmark dose modelling") is that the result of the benchmark dose (BMD) modelling exercise provides directly a description of the uncertainty: current dose-response modelling tools, such as the US EPA BMDS or the PROAST-based dose-response tool of EFSA calculate the confidence interval (CI) of the BMD, which is provided as the BMDL (5% one-sided lower limit of the BMD) and the BMDU (5% one-sided upper limit of the CI of the BMD). The more variable the observed data are or the fewer the number of animals per group is, the wider will the resulting confidence interval be, even if the central estimate (BMD) does not change. Therefore, using the BMDL as the POD allows considering the uncertainty inherent to the BMD in the assessment. Hence, we conclude that in a deterministic system for deriving OELs use of the BMDL is preferred over BMD. A probabilistic assessment allows using the whole distribution of the BMD (see examples in section 5).

The above-mentioned tools (BMDS and PROAST) have been further developed in recent times to consider an additional source of uncertainty: model uncertainty (uncertainty whether a single chosen model characterises best the dose-response). The newly introduced method of model averaging allows to use information from all models but gives more weight to those with better fits. Thus, instead of using the (potentially very conservative) lowest BMDL of the acceptable models, all models contribute to establishing of the CI of the BMD. Model averaging should be applied as a default procedure when performing dose-response modelling (see separate report "Benchmark dose modelling").

In a probabilistic assessment, instead of using the BMDL as the POD, the whole uncertainty distribution of the BMD can be used as a starting point of the assessment. As experience shows that the BMD distribution is typically skewed to the right side (a general behaviour for values with a boundary on the left side – benchmark doses are always >0), it is reasonable to assume a lognormal distribution for the BMD. Under the assumption of a lognormal distribution, a BMD distribution can be determined from BMDL and BMDU. This can be done, for example, with the EFSA MonteCarlo tool (see section 2.1.3). **Figure 5-1** and **Figure 5-3** show the resulting BMD(L) and distributions of the BMD for the two example substances used in this report.

In a separate report we performed dose-response modelling on ten example substances to demonstrate the applicability of deriving BMDL for various types of

substances, endpoints and studies (see report “Benchmark dose modelling – examples”). In the following table we compare the obtained BMDLs for these ten examples with the NOAELs (or LOAELs) determined by the study authors, evaluating bodies or – in case these are not available – by the authors of this report. This analysis shows that the position of NOAELs and LOAELs can vary substantially relative to the BMDL. For example, for divanadium pentoxide the obtained BMDL is very close to the LOAEC, whereas in the case of 3-MPCP the BMDL was lower than the LOAEL by a factor of 10. Where NOAELs were available these were 2- to 5-fold higher than the respective BMDLs. As a consequence of using LOAELs or NOAELs the resulting OELs differ in the protection achieved (due to different effect levels at the POD).

Table 4-1 Comparison of BMDL with NOAEL/LOAEL values for ten example substances

#	Substance	LOAEL	NOAEL	BMDL*
	Quantal data			
1	3-Monochloropropane-1,2-diol (3-MCPD)	1.97 mg/kg bw/d		0.19 mg/kg bw/d
2	Divanadium pentaoxide	0.28 mg/m³		0.23 mg/m³
3	4,4’-Methylene-bis-[2-chloroaniline] (MOCA)	9.4 mg/kg bw/d		2.91 mg/kg bw/d
4	Nitrilotriacetic acid (NTA) and its sodium salts	6.9 mg/kg bw/d		1.0 mg/kg bw/d
5	Benzoic acid	25 mg/m³	12.6 mg/m³	6.36 mg/m³
	Continuous data			
6	Nalidixic acid	8000 ppm food	4000 ppm food	1650 ppm food
7	1,1,2,2 Tetrachloroethane	40 mg/kg bw/d	20 mg/kg bw/d	49.3 mg/kg bw/d (at BMR 20% rel. liver weight change); 7.87 mg/kg bw/d (at BMR 7%)
8	N-octadecyl β-(3',5'-di-tert-butyl-4'-hydroxyphenyl) propionate (OBPP)	100 mg/kg bw/d	30 mg/kg bw/d	42.6 mg/kg bw/d (at BMR 20% rel. liver weight change); 6.86 mg/kg bw/d (at BMR 10%)
9	Tert-Butyl alcohol	180 mg/kg bw/d		47.7 mg/kg bw/d
10	Benzene	10.72 ppm-years	6.51 ppm-years	1.3 ppm-years

* $BMDL_{10}$ in case of quantal data, for BMRs for continuous endpoints see the report “Benchmark dose modelling – examples”)

In contrast to the BMD, the inherent uncertainty of a NOAEL is unknown. As discussed in the report on “Benchmark dose modelling”, the BMDL is expected to decrease when the number of animals per group were decreased (as the uncertainty increases). A NOAEL in such a situation might actually increase with decreasing numbers of animals, as the statistical power to show significance is reduced and, therefore, the next higher dose might be identified as NOAEL. Thus, whether a NOAEL is an adequate POD and close to the BMDL/BMD is expected to be case-specific and dependent on the substance-specific database and the quality of the selected key study.

4.2 Uncertainty of the NOAEL

As discussed above, NOAELs comprise inherent uncertainty, which is case-dependent and not reflected in the numerical value of the NOAEL. In the WHO/IPCS report on uncertainty in hazard assessment the authors tried to identify an uncertainty distribution for the NOAEL, using comparisons of the NOAEL and BMDL values (WHO 2014). Based on published comparisons between NOAELs and BMDLs for various types of data (continuous data and quantal data on developmental effects) they concluded that distributions assuming a 9-fold distance between the BMDL and BMDU would be appropriate. However, no data are available for quantal data from subchronic or chronic studies (on non-developmental effects).

No uncertainty distribution for NOAELs is derived here. The uncertainty associated with a NOAEL very much depends on the quality of the database in question. Whenever possible the BMD approach should be used to attain a description of the uncertainty of the POD. But if a NOAEL is used for an assessment, it should be kept in mind that its inherent uncertainty is not taken into consideration by applying assessment factors. Especially in case of poor data this should be given special consideration.

4.3 Difference between continuous and quantal data

In our report „Probabilistic hazard assessment" we already discussed differences between using continuous or quantal data as point of departure.

Continuous data typically comprise measured data for a continuous parameter represented as group means with standard deviation. A NOAEL or BMD determined from this kind of data gives the group average response. This is graphically shown in Figure 4-1 by locating the NOAEL/BMD in the middle of the density function representing the experimental animals' variability. In this case, the benchmark response chosen (or the type of effect observed at the LOAEL) defines the nature of the POD (and also that of the derived OEL).
For example, the nature of the POD would be different for the 2 hypothetical cases:

- BMR is a 2% increase in liver weight
- BMR is a 100% increase in serum glutamic-pyruvic transaminase activity

However, in both cases the observed dose-response data (and the BMD determined from them) represent average responses in the various dose groups with responses from animals showing strong effects and animals showing no or minimal effects averaged.

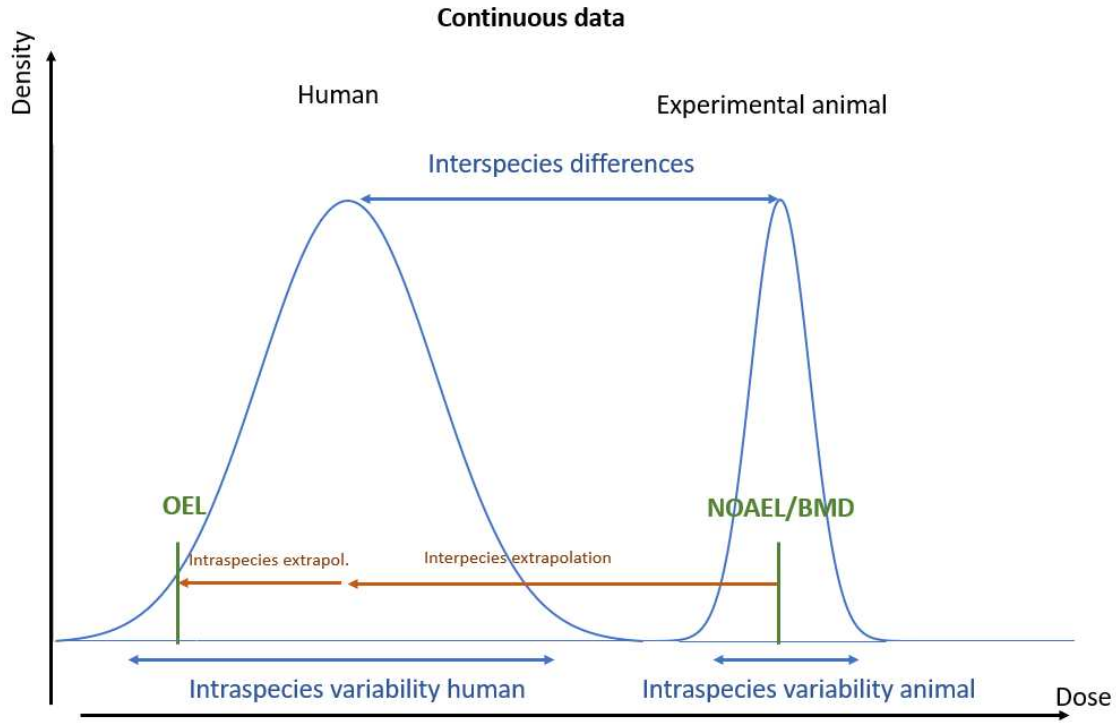


Figure 4-1 Continuous data: Relationship between inter- and intraspecies extrapolation

With quantal data (responses expressed as the percentage of responders, e.g. 10% incidence) the BMD or the NOAEL used as POD typically represents the dose leading to a low incidence of responders per group. This is graphically represented in Figure 4-2 by locating the NOAEL/BMD on the left side of the density distribution expressing the variability within the experimental animal species.

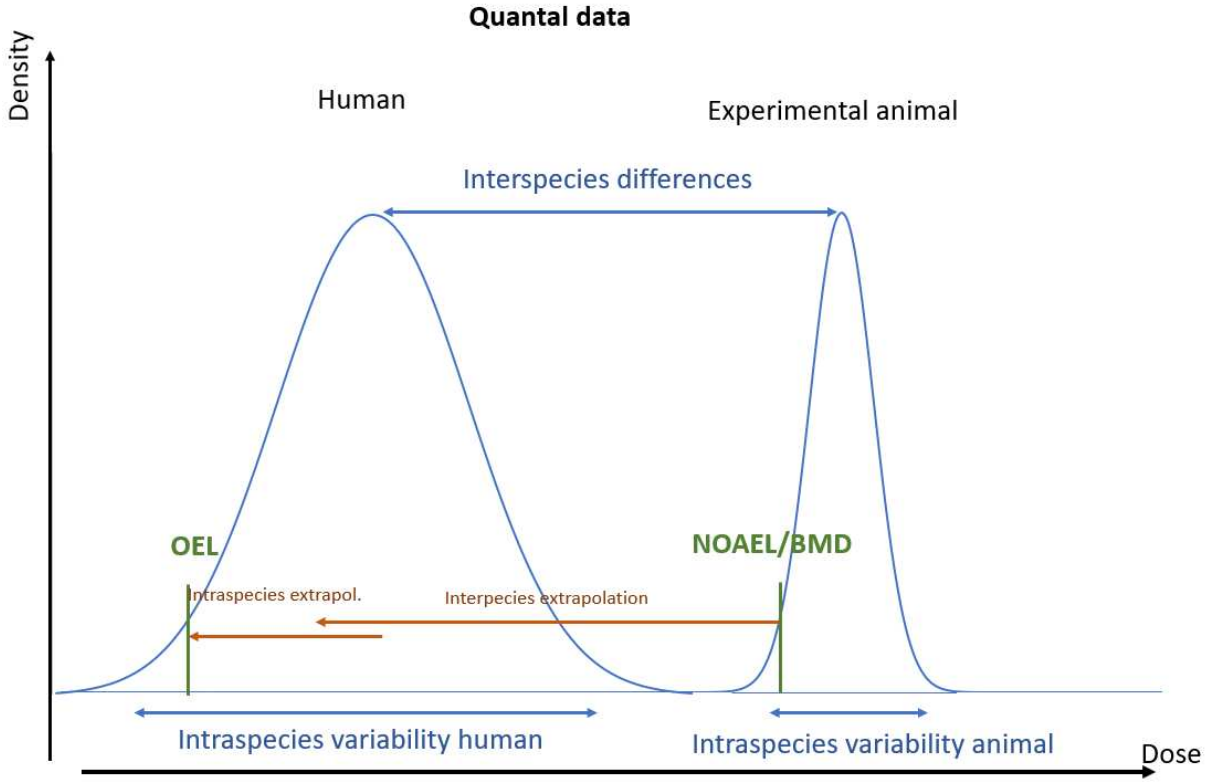


Figure 4-2 Quantal data: Relationship between inter- and intraspecies extrapolation

As the POD for quantal data includes this aspect of intraspecies extrapolation, the application of interspecies extrapolation for deriving OELs leads to accounting for intraspecies variability twice (which is not the case with continuous data). This is depicted by the overlapping arrows in Figure 4-2. However, as intraspecies variability of inbred experimental animals is expected to be much less compared to humans (represented by the two distributions of differing width) the resulting error might be small.

An alternative approach would consist in using the BMDL50 as the POD, as proposed in the WHO/IPCS report on uncertainty in hazard assessments (WHO 2014). However, it should also be noted that quantal data are to be considered a condensation (reduction) of continuous data, which contain less information and often measure the incidence of frank toxic effects (e.g. liver necrosis), which cannot be considered the borderline between adverse and non-adverse responses (which would be the ideal starting point for deriving OELs, which aim to avoid adverse effects in the target population).

4.4 Discussion

The nature of the POD defines to a large extent the nature of the resulting OEL. Using the BMDL as the POD provides for a more exact interpretation of the OEL. In addition, dose-response modelling allows to consider the uncertainty associated with the POD. In probabilistic approaches the uncertainty of the BMD can be described by a lognormal distribution informed by the BMDL and BMDU obtained from dose-response modelling.

The relationship of the NOAEL to the “true no adverse effect level” shows large variability, as shown in our comparison of BMDLs with NOAELs/LOAELs for ten examples. This uncertainty is typically not considered when NOAELs are used to derive OELs. The uncertainty is even higher with use of a LOAEL as POD.

The interpretation of a resulting OEL depends on the nature of the POD and the underlying effect size (for continuous and quantal data). As explained above, using a $BMD(L)_{10}$ as POD for quantal data contains an element of the interindividual variability of the experimental animals. On the other hand, a benchmark response for continuous data should be fixed at the borderline between non-adverse and adverse effects, whereas a POD derived from quantal data typically comprise already 5 to 10% incidence of the adverse effect considered.

5 Probabilistic assessment of two example substances

5.1 Introduction

For two example substances, deterministic OEL derivations according to relevant European and German OEL methodologies (REACH, RAC, AGS, MAK, EU BPR) are compared to a probabilistic assessment based on the same effects using a MC simulation with the distributions as presented in this report. For the probabilistic assessment, the POD is modelled as a lognormal distribution of the BMD, with BMDL and BMDU values as obtained from the dose-response modelling, while the POD for the deterministic assessment was the NOAEL/NOAEC. Although most of the methodologies support the use of a BMD as POD in theory, this is not yet common practice.

The substances were chosen based on timeliness (availability of a recent assessment or currently ongoing discussions), the availability of the dose-response data required for the benchmark dose modelling and on the type of effects which should provide the possibility to use a dose response model based on continuous data for one substance and on quantal data for the other substance.

The deterministic OELs are then compared to the probabilistic model to determine the probability that the OEL protects from the modelled effect. For example, if the probabilistic model determined a distribution of air concentrations with a 75th percentile of 10 mg/m³, then an OEL of 10 mg/m³ would correspond to a 25% probability of showing the modelled effect (as defined by the BMR) at or above the OEL with a certain incidence (1 and 5 % chosen in the examples) in the population of exposed workers.

5.2 Example substances

5.2.1 1,1,2,2-Tetrachloroethane

5.2.1.1 Database

1,1,2,2-Tetrachloroethane was chosen as the first example substance. This substance was recently re-evaluated by MAK Commission and an OEL of 14 mg/m³ was derived (Hartwig 2020). The critical study was the subchronic continuous feeding study by NTP with rats. In this study, rats were dosed with 1,1,2,2-Tetrachloroethane 7 days/week for 14 weeks via feed. The NOAEL was 20 mg/kg bw/d based on increased liver weight in both sexes and decreased sperm motility in males. When deriving their OEL, the MAK Commission considered an observed difference in absorption rates in rats: 95% oral absorption in rats versus an inhalation absorption rate in humans of 60%. Converting the dose to an air concentration yields a POD of 77.6 mg/m³, which includes correction for differences in the exposure scenario (more details in footnote 6). With a

factor of 2 for time extrapolation and a factor of 2 for inter- and intraspecies differences, an OEL of 14 mg/m³ was obtained (originally 19.4 mg/m³, but this value was rounded based on the concentration expressed in ppm).

5.2.1.2 Deterministic evaluations

POD modifications

The POD was modified according to the recommendations in the guidances to the investigated OEL frameworks. If compatible with the guidance, the following parameters were used for modifying the POD (which leads to a total adjustment factor of 3.88[6]):

- absorption differences were considered as evaluated by the MAK Commission: oral absorption rate in rats: 95%, inhalation absorption rate in humans: 60%.
- a five-day work week with a daily breathing volume of 10 m³ and a bodyweight of 70 kg was assumed compared to the continuous oral exposure in the rat study,

Application of a factor for allometric scaling, (4 in case of extrapolating from a rat study to humans), may occur either as a correction of the POD or as an assessment factor which is applied to the POD. In this example, the dose from the feeding study (in mg/kg bw/d) is converted to an air concentration. When converting from a dose, usually the allometric correction is included in this conversion in the form of breathing volume (which is proportional to caloric demand). Except of the EU BPR, all evaluated methodologies propose this, or are at least compatible with this approach. The BPR is special, because it does not use allometric correction but instead uses a large assessment factor (10) for intraspecies differences. For the rat to human extrapolation, this numerically corresponds to the combination of the allometric scaling with the factor for intraspecies variability. For this reason, the POD in **Table 5-4** is 4 times higher for the EU BPR than for the other methodologies. The footnotes in **Table 5-4** contain the calculation details for the POD modifications.

Assessment factors applied and resulting deterministic OELs

The evaluated methodologies propose different default factors for time extrapolation, interspecies differences and intraspecies variability to apply to the POD which are listed in **Table 5-4** as individual AF as well as a resulting total AF. As explained above, the EU BPR differs from the remaining methodologies in having a higher intraspecies AF which cancels out (in the case of a rat to human extrapolation with an allometric scaling factor of 4) with the higher POD. For the remaining methodologies, which start from the same POD, the differences in the total AF directly translate to the differences in the resulting OELs.

[6] Calculation of the POD adjustment: (absorption animal, oral / absorption human, inhalation) * (7 exposure days per week animals / 5 exposure days per week workers) * (allometric scaling) * (70 kg bodyweight /10 m³ breathing volume per workday) = (0.95/0.60) * (7/5) * (1/4) * (70/10) = 3.879

5.2.1.3 Probabilistic assessment

BMD distribution

For the dose response modelling, the increased relative liver weight in male and female rats (combined dataset) was used with a BMR of 20% increase in relative liver weight (see Table 5-1). This BMR was adopted from a respective decision of MAK Commission and AGS (unpublished minutes of Sub-Committee III of AGS, June 2016) (for details on the dose response modelling including the justification for the selected endpoint and BMR see the separate report “Benchmark dose modelling – examples”; see further below a discussion on the outcome using an alternative BMR).

With this BMR a BMDL of 49.3 mg/kg bw/d and a BMDU of 83.2 mg/kg bw/d were derived and, under assumption of a lognormal distribution, provide the 5th and 95th percentile for the distribution of the POD in our probabilistic hazard assessment. This distribution has the parameters **μ = 4.16 and σ = 0.16,** which is shown in Figure 5-1 and additional moments are given in Table 5-2.

Table 5-1 Data on relative liver weight in male and female rats (according to NTP (2004) used for benchmark dose modelling

Dose (mg/kg bw/d)	Relative liver weight (mean) (mg/g)	Relative liver weight (SEM) (mg/g)	N (# animals in group)	Sex
0	34.79	0.42	10	m
20	36.72	0.44	10	m
40	41.03	0.85	10	m
80	45.61	0.52	10	m
170	44.68	0.45	10	m
320	52.23	1.42	10	m
0	35.07	0.56	10	f
20	36.69	0.36	10	f
40	37.84	0.51	10	f
80	44.2	0.27	10	f
170	48.03	0.89	10	f
320	58.4	1.42	10	f

SEM: standard error of the mean; m: male; f: female

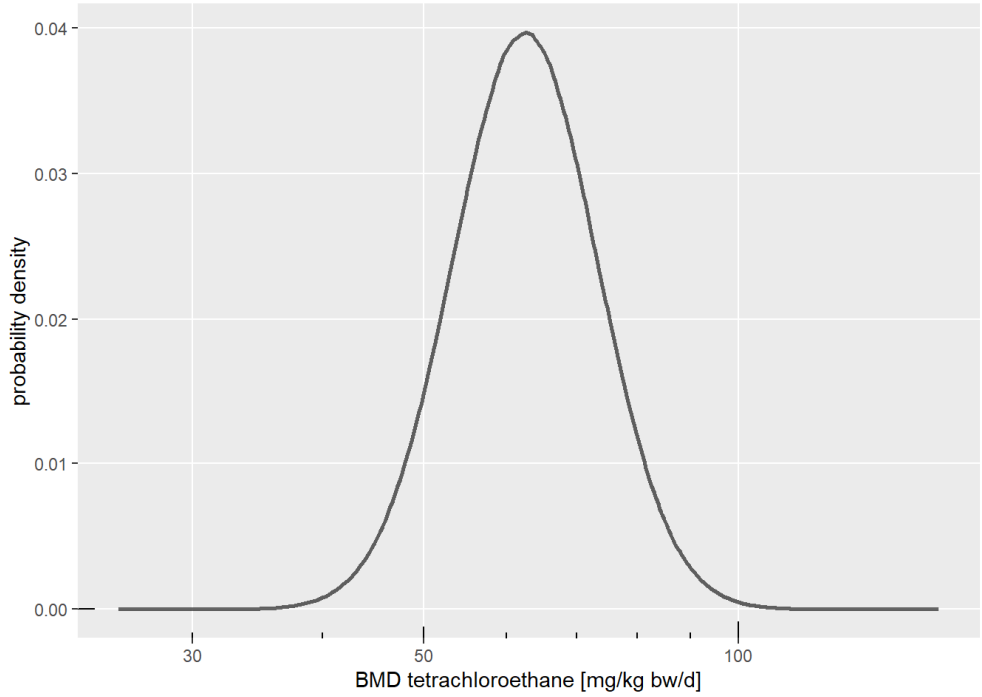


Figure 5-1 Probability density of distribution of the BMD for 1,1,2,2-tetrachloroethane, calculated based on the BMDL and BMDU of the model averaging using the EFSA PROAST tool.

Table 5-2 Parameters of the BMD (mg/kg bw/day) for 1,1,2,2-tetrachloroethane

Mean	**SD**	**GM**	**GSD**	**5%**	**Median**	**75%**	**95%**
64.87	10.39	64.05	1.17	49.30	64.05	71.30	83.21

In this example, the BMDL (49.3 mg/kg bw/d) is well above the NOAEL of 20 mg/kg d (the relative liver weight change at the identified LOAEL of 40 mg/kg bw/d is 18%, compared to controls; the modelling is in agreement with the dose-response data).

A BMR of 20% was set following conclusions of German committees for this endpoint based on Hall et al. (2012). This relative change in liver weight was defined by German committees as the borderline between pure adaptive responses and beginning liver toxicity (unpublished minutes of Sub-Committee III of AGS, June 2016). In an earlier approach to define benchmark response levels for a large set of continuous endpoints, Dekkers et al. consulted experts to obtain endpoint-specific recommendations for setting BMR values for continuous data (called “critical effect sizes” by the authors). For relative liver weights only one recommendation is reported for assessing human data (1st percentile), which cannot be applied to experimental animal data. For absolute liver weight proposals are either to use a relative liver weight change of 5 – 10% or two

standard deviations (variability in the control group) (Dekkers et al. 2001). Two standard deviations typically provide for a value largely outside the background variability observed in control animals. For the dataset under discussion here, setting the BMR at two standard deviations would lead to a 7% change in relative liver change. This BMR of 7% would lead to a BMDL of 7.9 mg/kg bw/d, a value substantially lower than the BMDL used above and lower than the NOAEL derived for this dataset. The rationale of using this BMR would be to avoid any clear-cut effects of the liver in the light of severe liver toxicity observed at higher doses. The observation of these large differences in the POD depending on the chosen BMR corroborates the importance of clearly justifying the BMR and the resulting point of departure. Implicitly, it also shows that using the arbitrary NOAEL (which depends heavily on the experimental setting) might have a high impact on the assessment, as it will only by chance correspond to a definite BMR set based on toxicological considerations.

Monte Carlo Simulation

The distribution of OELs is simulated with MC according to Equation 6.

$$\mathrm{OEL} = \frac{\mathrm{BMD} * \mathrm{POD\ adjustments}}{\mathrm{Time} * \mathrm{Interspecies} * TD_i * 10^{\log_{10}\mathrm{GSD} * z_{1-i}}} \qquad \text{(Equation 6)}$$

In this equation,

- 'BMD' refers to the distribution of the BMD, as obtained from the dose-response modelling (unit mg/kg bw/d)
- 'POD adjustment' is a constant factor, composed of the adjustments for differences in the exposure scenarios (see section 5.2.1.2: 3.879 kg bw/d * m^3)
- 'Time' is the distribution for time extrapolation subchronic to chronic (described in section 2.3)
- The 'Interspecies' distribution is described in section 2.4.3
- Intraspecies:
    - 'TD_i' is the distribution for the toxicodynamic fraction of the intraspecies variability with incidence i. The distributions for i = 1% and i = 5% are described in sections 2.5.2.1 and 2.5.2.2
    - '$\log_{10}$ GSD' is the distribution from section 2.5.1 which describes the variability in toxicokinetics. Finally, $z_{1\text{-}i}$ is the z-Score of the normal distribution with incidence i (the percentile with probability 1-i)

The probabilistic modelling was performed with the EFSA MonteCarlo tool (see section 2.1.3). The detailed protocol is attached in Annex 2. The following figure represents the results graphically, including a comparison with the deterministic evaluations.

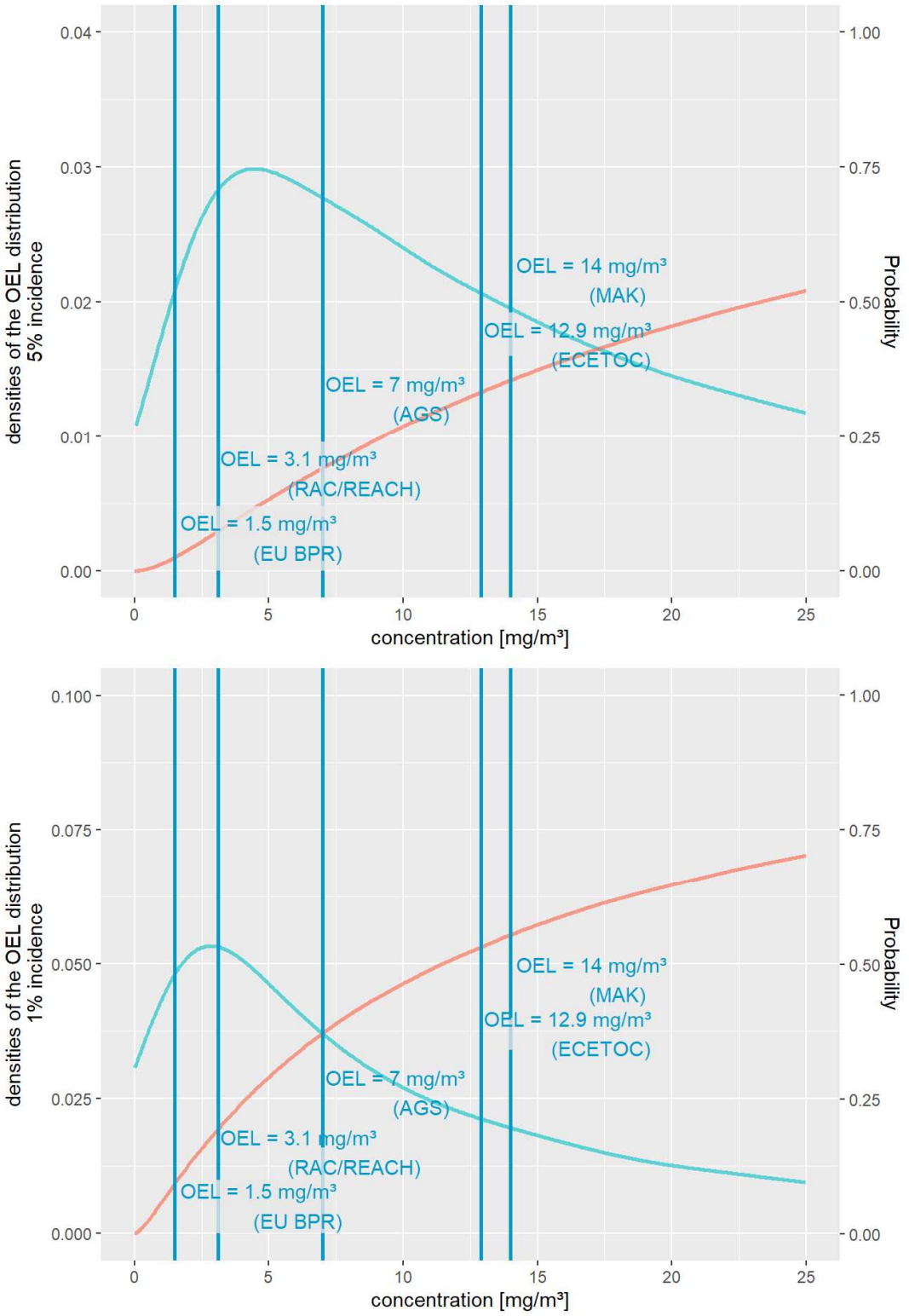


Figure 5-2 Example 1,1,2,2-tetrachloroethane: Probability density of the OEL (blue) and cumulative probability (red) distribution obtained by Monte Carlo simulation with the EFSA MonteCarlo tool (vertical lines represent the various deterministic OELs for comparison); top: incidence level 5%; bottom: incidence level 1%

Comparison of probabilistic with deterministic evaluations

The results of the MC simulations of Equation 6 are compared to the OELs from the deterministic evaluations. Each deterministic OEL corresponds to a probability p of the simulated distribution that the effects as described by the BMR actually occur. Increasing the OEL corresponds to an increase of p, but is less protective for the worker population, therefore, 1-p is the 'covered' probability by a certain OEL, that means the probability of this OEL to protect from the modelled effects in the simulation. These coverages range from 64.5 % (MAK) to 97.5 % (EU BPR) for the scenario with 5 % incidence and from 44.5 % (MAK) to 90.7 % (EU BPR) for the 1 % incidence scenario (**Table 5-4**).

Table 5-3 Parameters of the distribution of the probabilistic OEL (mg/m^3) for 1,1,2,2-tetrachloroethane

Incidence level	GM	GSD	5%	Median	75%	95%
1%	11.11	4.55	0.89	11.44	30.90	126.12
5%	22.95	3.96	2.36	23.29	58.11	214.97

Note: the distribution's percentiles give the probability p of experiencing effects as defined by BMR; coverage (probability that the OEL is providing the anticipated protection) is 1 – p.

Table 5-4 Modification of the POD for the study on 1,1,2,2-Tetrachloroethane (systemic effects, oral subchronic study with rats), resulting OEL and covered probability of this OEL according to the probabilistic model

Framework	POD	modified POD	proposed AF (time, inter, intra)	Total AF	OEL	Covered probability by OEL according to probabilistic model	
						5 % incidence	1 % incidence
RAC/REACH	20 mg/kg bw/day	77.6 mg/m³ *	2, 2.5, 5	25	3.1 mg/m³	92.7 %	80.7 %
AGS	20 mg/kg bw/day	77.6 mg/m³ **	2, 5, -	10	7 mg/m³	80.9 %	62.9 %
MAK	20 mg/kg bw/day	77.6 mg/m³ ***	2, 2, -	4	19.4 mg/m³, rounded to 14 mg/m³ (see text)	64.5 %	44.5 %
ECETOC	20 mg/kg bw/day	77.6 mg/m³ ****	2, 1, 3	6	12.9 mg/m³	66.7 %	46.9 %
EU BPR	20 mg/kg bw/day	310.4 mg/m³ *****	2, 10, 10	200	1.5 mg/m³	97.5 %	90.7 %

* RAC/REACH: REACH guidance R.8 provides a suggested default conversion from animal bodyweight doses to air concentrations for humans (* (1/0.38 m³/kg bw) * (6.7m³/10m³)). For comparison to modifications according to other guidances, here, the correction via default anthropometric data is used, which results in the same value: 20 mg/kg bw/day * (absorption animal, oral / absorption human, inhalation) * (7 exposure days per week animals / 5 exposure days per week workers) * (allometric scaling) * (70 kg bodyweight /10 m³ breathing volume per workday) = 20 mg/m³ * (0.95/0.60) * (7/5) * (1/4) * (70/10) = 77.6 mg/m³

** AGS: POD correction is the same as in the RAC/REACH scenario

*** MAK: POD correction is the same as in the RAC/REACH scenario

**** ECETOC: Guidance refers to R.8, same as RAC/REACH scenario

***** EU BPR: The BPR does not use allometric scaling, and instead uses a larger AF for interspecies differences. This leads to a four-times higher POD than for the other methodologies. Further, it does not provide defaults for anthropometric data which is why we use the defaults from RAC/REACH. Apart from allometric scaling, the correction is the same as in the RAC/REACH scenario.

5.2.2 Benzoic acid

5.2.2.1 Database

Benzoic acid serves as a case study for quantal data. In a recent evaluation by the MAK Commission, the critical effect for OEL derivation was interstitial inflammation and fibrosis of the lung in a 4-week inhalation study in rats. The rats were exposed for 6 hours on 5 days per week. Effects were observed at the lowest concentration in this study (25 mg/m^3), but a NOAEC of 12.6 mg/m³ was derived from a similar study, which did not reveal effects at the highest concentration tested (12.6 mg/m³). Starting from this NOAEC, a factor of 2 was applied for differences in the exposure scenario (please note the explanations below on this POD modification for local effects). With the total AF of 12, a rounded OEL of 0.5 mg/m³ was derived (Hartwig and MAK Commission 2018).

5.2.2.2 Deterministic evaluations

POD modifications

In the evaluation by the MAK Commission, the POD was corrected by division with the factor 2, which covers the differences of daily breathing volume during exposure between rats during the experiment and workers at the workplace. This factor is nearly equivalent to the calculation (6 hours / 8 hours) * (6.7 m³ / 10 m³) as e.g. proposed by ECHA guidance R.8 (ECHA 2012).

However, this is a local, irritative effect and ECHA guidance R.8 envisages a correction of the POD for exposure duration only, if data showing the total dose or time dependency of the effect is available. This is not the case and consequently in this case study no correction for exposure duration is performed for RAC/REACH and the guidances which recommend an equal procedure (ECETOC, EU BPR). We are aware that in practice, the burden of proof is often considered to be reversed, i.e. this correction, which represents the more conservative procedure, is only omitted if the concentration dependency of the effects has been shown. For the assessment of the AGS methodology, we assume that the MAK procedure (division of the POD by 2) is supported (**Table 5-8**). Consequently, the POD (after modifications) is 12.6 mg/m³ for RAC/REACH, ECETOC and EU BPR and 6.3 mg/m³ for AGS and MAK (**Table 5-8**).

Assessment factors applied and resulting deterministic OELs

The evaluated methodologies propose different default factors for time extrapolation, interspecies differences and intraspecies variability to apply to the POD. In **Table 5-8** the individual AF as well as a resulting total AF are listed. Although the POD differs at most by a factor of 2, the differences in the applied AF lead to quite a large range of resulting OELs from 0.08 mg/m³ (EU BPR) to 4.2 mg/m³ (ECETOC).

5.2.2.3 Probabilistic assessment

BMD distribution

The dose response modelling is based on the combined data for interstitial inflammation of the lung in males and females (**Table 5-5**). The incidence data are reported as categorical data. Only effects appearing not only locally (classified as "focal" or "multifocal"), but in a "generalized" form were considered and converted into incidences (all grades) (for more details see the separate report on "Benchmark dose modelling – examples").

Table 5-5 Data on interstitial inflammation ("generalized") of the lung in male and female rats (according to MAK commission) used for benchmark dose modelling

concentration (mg/m³)	effect # affected animals	n # animals in group	sex
0	0	10	m
25	3	10	m
250	4	10	m
1200	8	10	m
0	0	10	f
25	0	10	f
250	5	10	f
1200	9	10	f

m: males; f: females

The uncertainty distribution of the BMD, obtained for a benchmark response level of 10% incidence, by model averaging using the EFSA PROAST tool, had a 5th percentile of 6.37 mg/m³ (BMDL) and a 95th percentile of 92 mg/m³ (BMDU) which were used to derive a lognormal distribution. This lognormal distribution has the parameters **$\mu = 3.18$ and $\sigma = 0.81$** and is visualized in **Figure 5-3**. Additional moments of the distribution are shown in **Table 5-6**. This distribution provides the parameter POD for the probabilistic assessment according to Equation 6.

This rather broad BMD distribution encompasses with its BMDL and BMDU the LOAEC of 25 mg/m³ of the modelled study and the POD used for deriving the OEL. The broad distribution signals a relatively high uncertainty of the BMD distribution. This is also seen in the sensitivity analysis of the probabilistic modelling, where the BMD distribution is the largest source of uncertainty. The high uncertainty comes from the limited quality of the dose-response data in the relevant concentration range: at the lowest dose an incidence of 30% was observed in male animals, whereas in females none of the animals was affected at this dose. However, similar incidences were observed at the next higher concentration for both sexes. The low number of datapoints in the lower range, together with low numbers of animals per group lead to a high uncertainty.

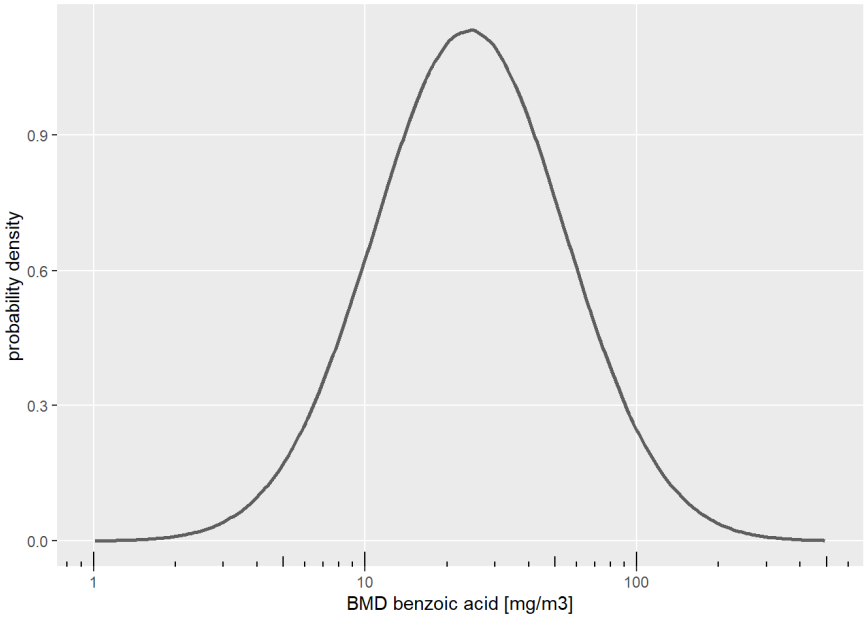


Figure 5-3 Probability density of distribution of the BMD for benzoic acid. Calculated based on the BMDL and BMDU of the model averaging using the EFSA PROAST tool.

Table 5-6 Parameters of the BMD (mg/m^3) for benzoic acid

Mean	SD	GM	GSD	5%	Median	75%	95%
33.64	32.50	24.19	2.25	6.37	24.19	41.83	92.00

Monte Carlo Simulation and comparison with deterministic OELs

The probabilistic modelling of the OEL distribution is performed with the EFSA MonteCarlo tool analogous to Equation 6 in section 5.2.1.3. Our own data evaluation indicated that also local effects in repeated dose studies show a time dependency. In the absence of substance-specific data that proves that the effects are driven by concentration only, an adjustment of the POD for differences in the exposure situation is warranted. Consequently, we adjusted the POD for the probabilistic modelling with the factor 0.5025 ((6 hours / 8 hours daily exposure) * (6.7 m³ respiratory volume at rest/ 10 m³ respiratory volume workers)).

This simulation is compared to the deterministic OELs. The coverages are generally higher than for the example substance 1,1,2,2-tetrachloroethane and range from 17% (ECETOC) to 93% (BPR) for the scenario with 5% incidence. The scenario for 1%

incidence produced coverages slightly lower with a range from 9% (ECETOC) to 83% (BPR).

Table 5-7 Parameters of the resulting distribution of the probabilistic OEL (mg/m^3) for benzoic acid

incidence level	GM	GSD	5%	Median	75%	95%
1%	0.41	5.73	0.02	0.42	1.36	6.93
5%	0.86	5.09	0.06	0.86	2.58	12.1

Note: the distribution's percentiles give the probability p of experiencing effects as defined by BMR; coverage (probability that the OEL is providing the anticipated protection) is 1 – p.

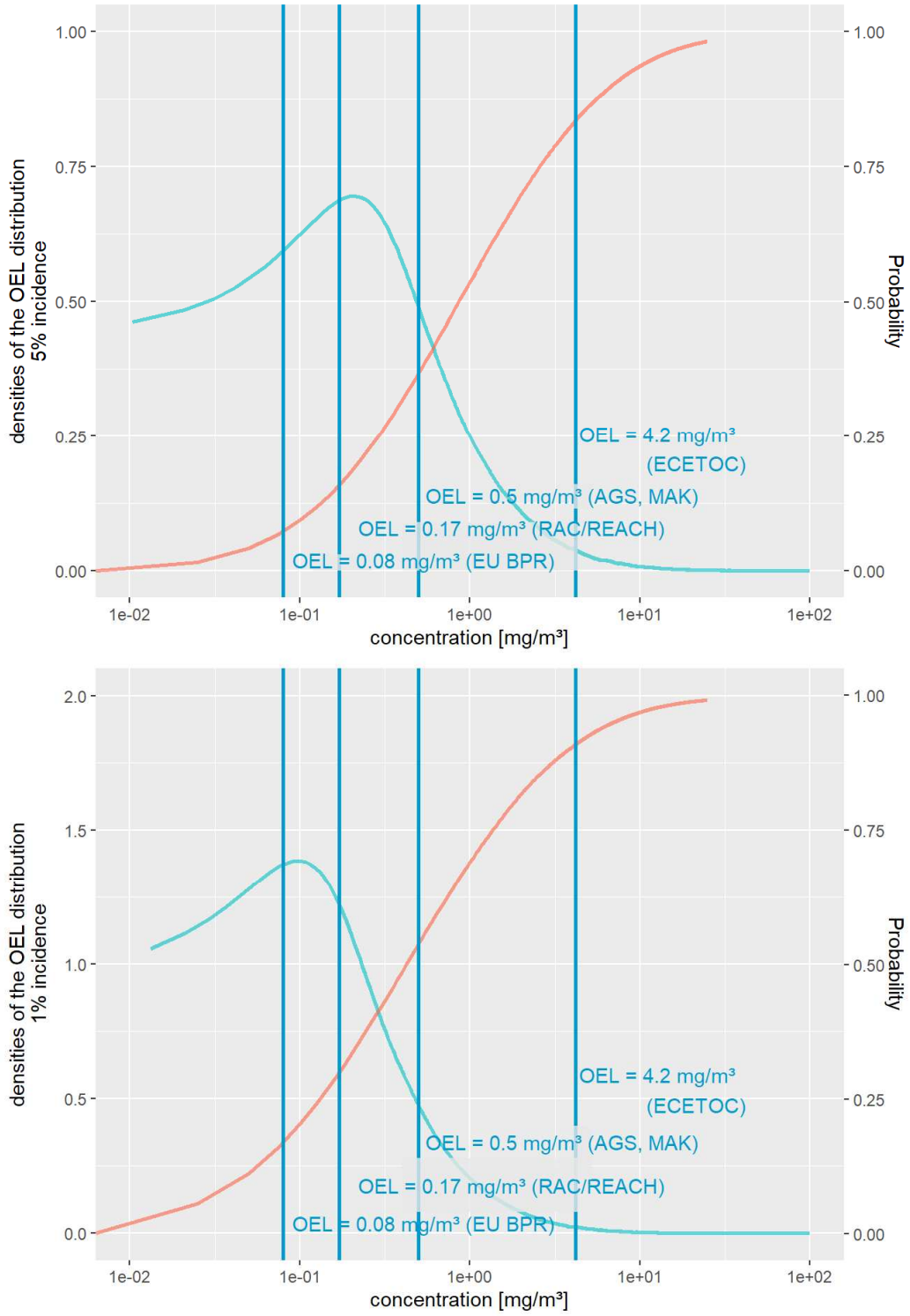


Figure 5-4 Example benzoic acid: Probability density of the OEL (blue) and cumulative probability (red) distribution obtained by Monte Carlo simulation with the EFSA MonteCarlo tool (vertical lines represent the various deterministic OELs for comparison); top: incidence level 5%; botton: incidence level 1%. Because of the high value for the ECETOC OEL, the x-axis is given as a logarithmic scale.

Table 5-8 Modification of the POD for the study benzoic acid (local effects in the lung, subacute inhalation study), resulting OEL and covered probability of this OEL according to the probabilistic model

Framework	**POD**	**modified POD**	**proposed AF (time, inter, intra)**	**Total AF**	**OEL**	**Covered probability by OEL according to probabilistic model**	
						5% incidence	**1% incidence**
RAC/REACH	12.6 mg/m³	12.6 mg/m³	6, 2.5, 5	75	0.17 mg/m³	84.1 %	70.1 %
AGS *	12.6 mg/m³	6.3 mg/m³	6, 2, -	12	0.5 mg/m³	63.3 %	46.1 %
MAK	12.6 mg/m³	6.3 mg/m³	6, 2, -	12	0.5 mg/m³	63.3 %	46.1 %
ECETOC	12.6 mg/m³	12.6 mg/m³	1, 1, 3	3	4.2 mg/m³	16.4 %	9.0 %
EU BPR	12.6 mg/m³	12.6 mg/m³	6, 2.5, 10	150	0.08 mg/m³	92.7 %	83.1 %

* Formally, with a default assessment factor of 30 (6 for time extrapolation and 5 for combined inter- and intraspecies extrapolation, the resulting OEL would be 0.2 mg/m³; however, assuming a lower inter- and intraspecies variability due to the acidic nature of the substance and a flat dose-response relationship the proposal by the MAK Commission was adopted by AGS (unpublished minutes of Sub-Committee III of AGS, September 2017)

5.3 Discussion

Probabilistic modelling of the two example substances results in distributions, which yield a range of different coverage values when compared to deterministic OELs.

For 1,1,2,2-tetrachloroethane coverage at the 5% incidence level is 65 - 98% and 45 - 91% at the 1% incidence level. These values are similar to those achieved by comparing distributions for the combined assessment factors with the respective distributions (see section 3.5; however note that in section 3.5 time extrapolation subacute to chronic was used).

The example also showed that the selection of the BMR for the continuous endpoint (relative liver weight) had a significant impact on the outcome. The dose-response data used for 1,2,2-tetrachloroethane are shown in Figure 5-5. The data show a rather linear behaviour at low doses. Whether a dose would be identified as a NOAEL (i.e. a dose with a change, which proves to be statistically significant), would depend on the data variability and the number of animals used per group. In contrast, if dose-response modelling is applied, the chosen BMR defines where the BMDL is located. In this case of a modest linear increase in effect size over a broad dose range selection of the BMR has a large impact on the numerical value of the BMDL.

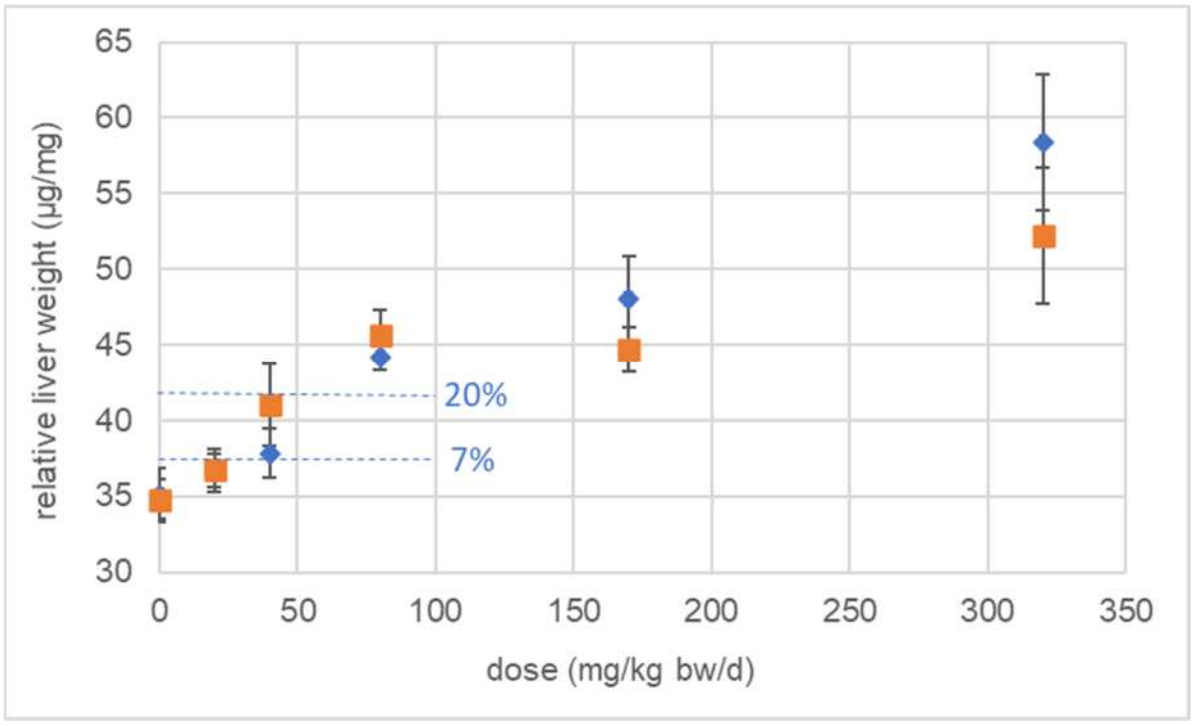


Figure 5-5 1,1,2,2-Tetrachloroethane: Dose-response data for liver weight in male (squares) and female (diamonds) rats (horizontal lines: low: BMR 7%, corresponding to 2 SD; high: BMR 20%, set by German committees based on toxicological considerations)

For benzoic acid, the achieved probabilities (coverage) were 84 and 93% for RAC/REACH and BPR, respectively, at the 5% incidence level, which is again similar to the coverage achieved when comparing assessment factors with the respective distributions (see section 3.5). For the MAK Commission's OEL, identical to the value set by AGS, the coverage is 63%. For ECETOC it is only 16%, as for local effects in the respiratory tract ECETOC foresees low or no assessment factors at all. Coverages at the 1% incidence level are lower, as can be seen in **Table 5-8**.

Adaption of the POD for differences in exposure conditions between the experimental study and the workplace situation is treated differently between AGS/MAK Commission and the ECHA Guidance R.8 in this case of local effects in the lower respiratory tract. This effected to reduce the differences between AGS/MAK Commission on the one hand and the other organisations on the other hand.

The discussion on the example 1,1,2,2-tetrachloroethane showed that the choice of the POD might have a large impact on the resulting distribution/OELs. Further, as expected, the differences in the size of the assessment factors used are reflected in the coverage achieved.

6 Discussion of existing methodologies and their protection goals

6.1 Discussion of protection goals associated with currently used methodologies

6.1.1 Principal considerations

In this project for the steps
- Time extrapolation
- Interspecies extrapolation
- Intraspecies extrapolation

new databases were created and discussed in the light of existing knowledge. On this basis, distributions were developed, which can be combined by probabilistic modelling and thus compared to currently used assessment factors. The obtained distributions show a high concordance with distributions proposed in a WHO/IPCS report (WHO 2014), but are mostly narrower, indicating less uncertainty, respectively intraspecies variability.

As input, the probabilistic assessment requires information on the percentage of the exposed population (workers) the assessment tends to protect. As a matter of principle, the uncertainty rises at the lower and upper end of a distribution. Because of this non-linear behaviour high factors are required to reach a very high coverage of the exposed population (e.g. beyond 99% probability).

We cannot pre-empt the regulatory decision on the percentage of populations to be covered by an assessment. For comparing the probabilistic assessment with current methodologies, we use two different distributions for intraspecies variability
- One covering 95% of the exposed population (5% incidence level)
- One covering 99% of the exposed population (1% incidence level).

For each of these two settings the probabilistic assessment results in a distribution describing the uncertainty of the OEL. This distribution can be compared to assessments based on currently used methodologies and a conclusion can be drawn on the probability of a current assessment achieving its protection goal.

If the probability with which a probabilistic assessment should achieve its protection goal (for example, 75, 90 or 95%) is determined, then a single deterministic value can be derived from the probabilistic evaluation.

6.1.2 Comparison of individual extrapolation steps

In section 3 we compared the distributions obtained for each extrapolation step with the assessment factors in use.

For time extrapolation it was noted that for subacute to chronic extrapolation the assessment factor of 6 results in a probability of 68%, whereas the assessment factor of 2 for subchronic to chronic results in a probability of only 36%. This indicates that the subchronic to chronic factor should be higher to achieve a similar coverage as the subacute to chronic factor.

Our evaluation supports the concept of caloric demand scaling (although with the databases evaluated only a comparison between rats and mice was possible). When comparing the distribution for remaining uncertainties (after caloric demand scaling this distribution centres around 1, which means that experimental animals can be more or less susceptible than humans) with the assessment factor of 2.5 used in REACH and by RAC for OEL setting, then a relatively high probability of 88% results. The same coverage is achieved by the factor of 10 used in the PPP and BPR frameworks (without caloric demand scaling), if the POD is based on a rat study. For smaller animals, the probability is expected to be lower, for larger animals higher. As expected, without an assessment factor for remaining uncertainties (default value of 1 as proposed by ECETOC) the probability is about 50%.

Intraspecies extrapolation is the step with the poorest empirical database so far. We developed separate distributions for toxicokinetic as well as for toxicodynamic differences in the adult population and combined these to a single distribution for intraspecies extrapolation. This was done, as described above, for the two protection goals coverage of 95 and 99% of the exposed population. At the 95% level both the factors 5 and 10 achieved reasonable probabilities (73 and 95%). However, at the 99% level none of the factors in use achieved probabilities above 66%. This makes evident that a clear definition of the protection goals an OEL is aiming at is required.

The combined inter- and intraspecies factor of 5 as used by AGS covers 35% of the distribution in case of 99% of population is included and 62%, if 95% are included. The factor appears to be too low to achieve a high probability, but again the definition of the population to be included is critical.

6.1.3 Combination of assessment factors

When combining the distributions for time, inter- and intraspecies extrapolation and comparing the resulting distribution to the product of the individual assessment factors, the following sequence (from high to low probability) results:

BPR > RAC/REACH > AGS > ECETOC.

Further observations are:

- The PPP framework does not provide default values for all steps but can be assumed to be similar to BPR: 86% probability at the 99% population coverage level for the combined assessment factor of 600 (assumed starting point: subacute rat study).
- Assessment factors by RAC/REACH at this level lead to 73% probability of the combined assessment factor of 300.

- The framework of MAK Commission in Germany, although otherwise similar to AGS, does not contain the combined assessment factor of 5 and is therefore assumed to result in lower probabilities compared to AGS (for which a 51% probability was obtained at the 99% population coverage level with the combined assessment factor of 120).
- ECETOC proposes lower assessment factors compared to the others in most cases and, hence, results in low probabilities: 38% at the 99% population coverage level for the combined assessment factor of 72. For local effects ECETOC assumes strict concentration-dependency, resulting in even lower probabilities (see below, section 6.2.1).

Regarding results for distributions including 95% of the population see section 3.5 and **Table 3-6**: probabilities ranged from 95% probability (BPR) over 88% for REACH/RAC to 70% for AGS and 57% for ECETOC.

Overall, only the extrapolation factors used in the BPR framework result in a high probability (86%) of covering a large fraction (99%) of the targeted population. The major difference to REACH/RAC and AGS consists in the higher assessment factor for intraspecies extrapolation. Indeed, the evaluation of data for the adult healthy human population revealed that interindividual variability is high and may require higher assessment factors than currently used in many frameworks.

6.1.4 Comparison with the probabilistic approach

In addition to the uncertainty resulting from individual extrapolation steps further uncertainty is added to the assessment through the POD. Using a distribution of the BMD, which can be derived directly from the dose-response modelling, allows for considering the uncertainty of the BMD. In contrast, the unknown uncertainty of the NOAEL is typically ignored in current assessments. Further, in section 4.2 we discussed that a NOAEL can vary in its conservativism from case to case, depending on data quality, type of endpoint assessed and other factors.

How the individual assessment steps are interrelated and how a probabilistic assessment might deviate from a deterministic setting of an OEL is illustrated in section 5 with two examples. These examples also show that the POD is a very critical input parameter. For 1,1,2,2-tetrachloroethane it was exemplified that choosing a specific BMR for continuous endpoints might have a large impact on the numerical value of the BMDL and that the effect size at the NOAEL might vary substantially from a POD defined by toxicological criteria. For both examples the coverage (i.e. the probability that the OEL achieves the intended protection goal) provided by the respective OELs was in a similar range as the probabilities reported above (section 3.5) for the combined extrapolation steps.

In general, the large variation in assessment factors used in the different methodologies are reflected in corresponding differences in the coverage achieved by the respective OELs.

Probabilistic models were so far developed mostly for characterising hazards for the general population, not for workplace exposures. As discussed in section 2.6 distributions obtained with our databases are comparable with distributions proposed in a report by WHO (2014). Few data exist to characterize additional inter-individual variability introduced by consideration of susceptible groups such as children, elderly or sick people. Therefore, (the small) differences between the distributions obtained are largely caused by differences in the databases used.

The only probabilistic approach for deriving OELs we are aware of was published by our group several years ago (Schneider et al. 2004; Schneider et al. 2006). A comparison of the distributions used by Schneider et al. (2006) with those newly developed here is shown in section 2.6. The databases prepared in this project, especially those for time extrapolation and intraspecies extrapolation, can be considered an improved basis for deriving the distributions.

6.2 Methods for assessing local effects

6.2.1 Assessment factors for effects in the respiratory tract

In most frameworks default assessment factors for local effects in the respiratory tract are identical to those used for systemic effects. But defaults for time extrapolation may be reduced if it can be shown that effects are not increasing with increasing exposure time (ECHA 2012). The default position, however, is to use the same time extrapolation factors as for systemic effects.

Interestingly, no adaptation of the POD for differences in exposure time per day and for physical activity is foreseen in the Guidance document R.8 in case of local effects in the respiratory tract, unless there is evidence for a dose and/or time dependency: "*However, these modifications would only apply when there is evidence that the inhaled dose or duration of exposure, and not the concentration, drive the appearance of the effect*" (ECHA 2012). This difference in default positions for local effects for time extrapolation and for modifying the POD can be considered an inconsistency in the guidance document.

Our analysis of data from repeated dose studies with different exposure durations (subacute to chronic) support the conclusion that also local effects in the respiratory tract become more severe with increasing exposure time. A common default position of correcting for exposure time in both time extrapolation and POD modification would be in line with this evidence.

In contrast to other organisations ECETOC assumes as default position that local effects in the respiratory tract are exclusively concentration dependent. Therefore, ECETOC's default factor for time extrapolation is 1.This position was not based on empirical data, but on the theoretical assumption of concentration dependency for effects at the port of entry (ECETOC 2003, 2010).

Our analysis gave rise to the conclusion that the same assessment factors should be used for systemic effects and local effects in the respiratory tract. We already pointed out that the observed differences regarding interindividual differences in toxicokinetics after inhalation exposure warrant further investigations.

6.2.2 Inter- and Intraspecies variability in case of particles assessed by the HEC procedure

The procedure to derive a “Human equivalent concentration” (HEC) for inhalation exposure to particles was discussed in detail in a separate report of this project (“Human equivalent concentration and kinetic modelling of aerosols in the lower respiratory tract”). It includes application of a dosimetry model (MPPD model) and aims at correcting the POD (the inhalation concentration in an experimental animal inhalation study) for interspecies differences in deposition and clearance of particles.

When applying the HEC procedure in the case of particulate exposure, it can be argued that an assessment factor for potential differences between animals and humans (a factor of 2.5 is used by many organisations) can be reduced, as only uptake from the lung into the blood circulations and potential toxicodynamic differences have to be accounted for.

As part of the German procedure to derive OELs, for inter- and intraspecies extrapolation a combined factor of 5 is used (AGS 2010) (see also section 3.4). Currently there is no guidance on how to use the HEC procedure **in combination** with this combined factor for inter- and intraspecies extrapolation. It has been proposed occasionally to reduce the aggregated variability factor of 5. However, the analysis of the HEC procedure concluded that the procedure itself is associated with a high uncertainty (with regard to selecting the most appropriate normalisation procedure and adequate consideration of species differences in clearance, to mention only the two most important questions), which exceeds the factors 2.5 or 5 by large.

In our separate report on “Intraspecies extrapolation” several studies are cited which point to relevant interindividual differences with regard to particle deposition and clearance in humans. Further, the analysis in this report showed that a combined inter- and intraspecies factor of 5 leads to a rather low coverage of 34.5% of the distribution in case of an incidence level of 1% and 61.9% for a 5% incidence level.

Overall, even when the HEC procedure would accomplish to reduce the uncertainty in interspecies extrapolation of particles, the combined factor of 5 in the German methodology rather appears to provide a low level of coverage.

Due to the high inherent uncertainty of the HEC procedure we propose to use the same distribution or assessment factors to account for uncertainties in interspecies extrapolation in case of systemic effects and in case of particulates assessed by the HEC procedure.

6.2.3 Evaluation of sensory irritation

Brüning et al. (2014) proposed a scheme for deriving OELs based on sensory irritation as observed in human studies and animal experiments. Sensory irritation is defined as caused by an "interaction of local irritants with receptors of the nervous system (e.g., trigeminal nerve endings)", followed by a "cascade of reflexes and defence mechanisms (e.g., eyeblinks, coughing)" (Brüning et al. 2014). The concept of Brüning and colleagues suggests using the following default values for deriving an OEL based on sensory irritating effects observed in an experimental animal study:

Assessment factors for	
Time extrapolation:	6 (subacute – chronic) and 2 (subchronic – chronic)
Interspecies extrapolation:	3 (might be lowered to 2 in case of effects in the olfactory epithelium)
Intraspecies extrapolation:	1

With regard to inter-individual variability the authors discussed several studies investigating potential differences in susceptibility between groups of chemosensitive or allergic individuals and groups of normal healthy subjects and concluded "*that an intraspecies default factor is not necessary if OELs are derived from human sensory NOAECs since it is based on a controlled human exposure study assessing especially sensitive and objectively verifiable effects"*. Indeed, there is little evidence that asthmatics or atopic individuals are substantially more susceptible to sensory irritation than uncompromised adults (Nielsen and Wolkoff 2017). However, there are several reports pointing out that women might be more sensitive than men (Pacharra et al. 2016; Shusterman and Balmes 1997; Sucker et al. 2019). Further, age is discussed as a parameter influencing susceptibility. These data are discussed in more detail in our report on "Intraspecies extrapolation".

In our project we did not specifically aim at addressing the endpoint of sensory irritation and did not create databases specific for this endpoint. The findings regarding potential sex-related differences in susceptibility warrant to investigate further which reasons for inter-individual variability exist and whether a default factor of 1 is adequate.

6.3 Recommendations

Comparing currently used assessment factors with the distributions derived in this report allows to evaluate the probability with which a certain factor is expected to be adequate when applied in an OEL derivation. For example, the time extrapolation factor of 6 for subacute to chronic extrapolation provides a probability of 68%.

The other way round, by defining their protection goals, regulatory bodies can derive assessment factors in a consistent and transparent way. For example, if the protection goal is to achieve a 50% coverage for time extrapolation, assessment factors could be set at the medians of the respective distributions. For example, for subacute to chronic time extrapolation, an AF based on the median of the distribution would be 3.7, an AF based on the 75th percentile would be 7.5. For subchronic to chronic extrapolation the respective median would be 2.8 and the 75th percentile 5.5.

For interspecies extrapolation (in combination with using allometric scaling factors based on caloric demand scaling) the median would be 1 (as species are considered equally susceptible on average after application of allometric scaling) and the 75th percentile would be 1.7.

In case the HEC procedure is applied to consider interspecies differences in deposition and clearance of particles in the lower respiratory tract, as explained above, due to the high inherent uncertainty of the HEC procedure we propose to use the same distribution or assessment factors as for systemic effects.

For intraspecies variability risk assessors further need to decide on the fraction of the exposed population, for which the OEL cannot guarantee absence of adverse effects. Note that inclusion of 100% of the population would require unjustifiable low OELs, because the uncertainty at the outer limits of the distributions increases more than linear.

If intraspecies extrapolation aims at including 95% of the population (5% incidence level), then an

- assessment factor of 3.6 is required for achieving a 50% probability
- and a factor 5.2 for 75% probability.

If intraspecies extrapolation aims at including 99% of the population (1% incidence level), then an

- assessment factor of 7.3 is required for achieving a 50% probability
- and a factor 12.5 for 75% probability.

Combining several factors at a high probability level (e.g. above 75%) would lead to a very conservative OEL setting, whereas using assessment factors all derived from the median of the distributions lead to an OEL expected to cover about 50% of cases only.

Probabilistic methods allow to combine all distributions and to derive a probability function for the OEL. A deterministic OEL can then be obtained by defining its probability, i.e. choosing 1 point from the probability function.

For example, **Figure 6-1** shows the outcome of the probabilistic modelling for 1,1,2,2-tetrachloroethane (example 1, section 5.2.1). To obtain an OEL (definite single value) the incidence level has to be chosen (in this example set to 5%) as well as the coverage (probability that the factor provides sufficient protection) Choosing a coverage of 75% (corresponding to the 25th percentile of the distribution, see y-axis of the figure) would result in a value of 9.2 mg/m^3. To achieve 90% coverage (corresponding to the 10th percentile) a lower value of 3.9 mg/m^3 is required.

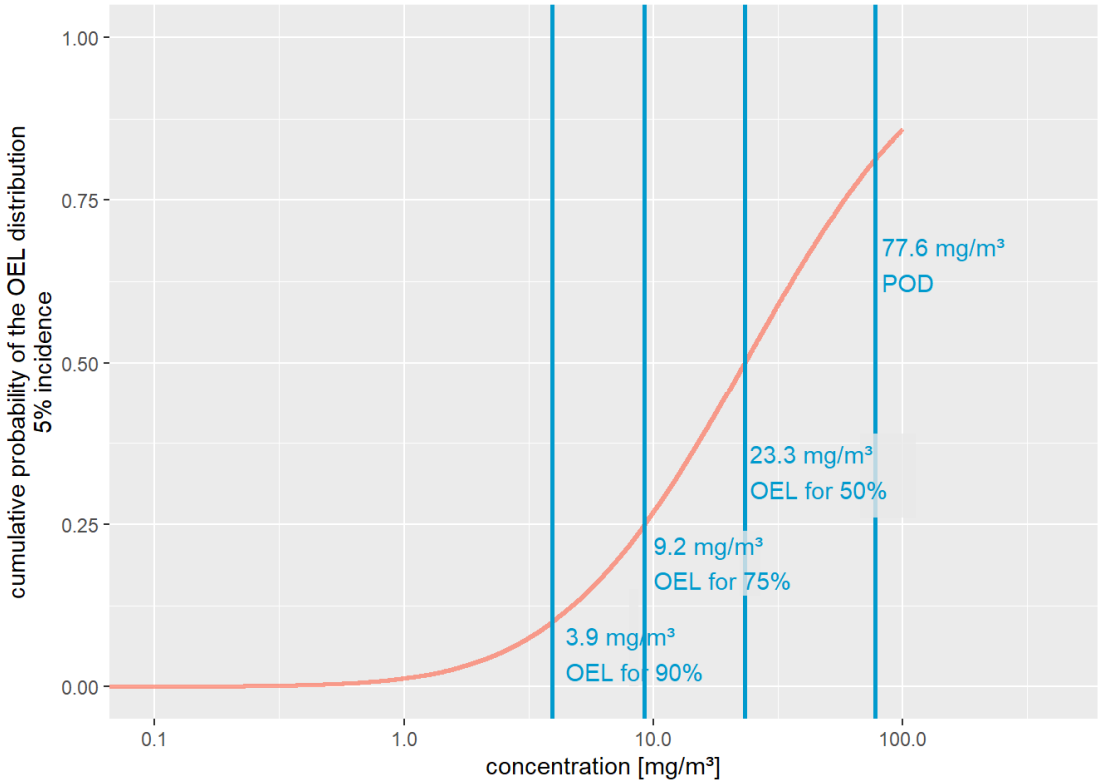


Figure 6-1 Example 1,1,2,2-tetrachloroethane: Cumulative probability distribution obtained by Monte Carlo simulation with the EFSA MonteCarlo tool for the incidence level 5% (dose in logarithmic scale)

Clearly defining the covered/not covered fraction of the population and the probability, with which the OEL aims to be safe, provides a transparent rationale for choosing assessment factors, allows to explain potential differences between methodologies and thus contributes to harmonise approaches.

Recommendation 1:

All OEL derivation frameworks should clearly define their protection goals by stating:

- The fraction of the exposed population covered by the OEL
- The probability with which they intend to provide protection from adverse effects (as defined by the POD)

As a NOAEL or LOAEL

- does not provide information on the level (incidence, severity) of effects at this dose level,
- does not allow to consider the differences between quantal and continuous data at the POD and thus does not allow to define the OEL in this regard

- does not provide information on the uncertainty associated

they should be used as POD only in cases in which benchmark dose modelling is not applicable.

Recommendation 2:

Benchmark dose modelling should be used as the default procedure to derive a POD

Probabilistic models allow to investigate the effect of combining deterministic assessment factors and the protection goals achieved by their combined use. Although it is unlikely that they will be used as a default procedure for deriving OELs soon, they can be used to adjust and standardize deterministic procedures.

Recommendation 3:

Probabilistic models should be further developed and used for benchmarking against deterministic methodologies to test them

For some of the distributions developed here further improvements seem possible. For the first time a database for toxicokinetic intraspecies variability based on inhalation data was established. These data hint on a lower variability compared to oral data. Conflicting results were obtained from example substances with inhalation exposure and toxicokinetic and toxicodynamic influences.

No other significant differences, regarding parameters sex, route, endpoints (local versus systemic effects), target organs or chemical structure became evident in our evaluations.

Recommendation 4:

Increasing and improving the database on inter-individual variability in human toxicokinetic inhalation studies might allow to establish route-specific distributions for intraspecies variability.

7 Conclusions

The analysis of existing frameworks for deriving OELs showed that indeed relevant differences exist between them. A major difference is in the size of assessment factors chosen and the comparison with the derived distributions for the individual extrapolation steps showed that these differences result in large differences in the coverage achieved (i.e. the probability that the OEL is protective enough).

Methodologies to derive OELs should define their protection goals regarding the percentage of the population included and the coverage (probability) they aim to achieve. This would be a major step forward towards increased transparency (as it facilitates comparing assessment factors and understanding differences) and possibilities for harmonisation.

Probabilistic modelling can have an important role here. It is not anticipated to become a standard tool to derive OELs soon, but it can be valuable to explain the coaction of extrapolation steps and the overall protection achieved by individual frameworks.

References

Abdo, N., Xia, M., Brown, C.C., Kosyk, O., Huang, R., Sakamuru, S., Zhou, Y.H., Jack, J.R., Gallins, P., Xia, K., Li, Y., Chiu, W.A., Motsinger-Reif, A.A., Austin, C.P., Tice, R.R., Rusyn, I., Wright, F.A. (2015). Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 genomes high-throughput screening study. Environmental Health Perspectives 123, 458-466.

AGS, Ausschuss für Gefahrstoffe. (2010). Bekanntmachung zu Gefahrstoffen. Kriterien zur Ableitung von Arbeitsplatzgrenzwerten. BekGS 901. Das Gemeinsame Ministerialblatt 61 (32), 691-696. http://www.baua.de/cln_135/de/Themen-von-A-Z/Gefahrstoffe/TRGS/Bekanntmachung-901.html.

Belgorodski, N., Greiner, M., Tolksdorf, K., Schueller, K., Flor, M., Göhring, L. (2017). Package 'rriskDistributions', Online: https://cran.r-project.org/web/packages/rriskDistributions/rriskDistributions.pdf (08.11. 2018).

Bokkers, B.G.H., Slob, W. (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. Critical Reviews in Toxicology 37, 355-373.

Brüning, T., Bartsch, R., Bolt, H.M., Desel, H., Drexler, H., Gundert-Remy, U., Hartwig, A., Jäckh, R., Leibold, E., Pallapies, D., Rettenmeier, A.W., Schlüter, G., Stropp, G., Sucker, K., Triebig, G., Westphal, G., van Thriel, C. (2014). Sensory irritation as a basis for setting occupational exposure limits. Archives of Toxicology 88, 1855-1879.

Dekkers, S., de Heer, C., Rennen, M.A.J. (2001). Critical effect sizes in toxicological risk assessment: a comprehensive and critical evaluation. Environmental Toxicology and Pharmacology 10, 33-52.

ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals. (2003). Technical Report No. 86. Derivation of Assessment Factors for Human Health Risk Assessment, Brussels, Belgium.

ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals. (2010). Technical Report No. 110. Guidance on Assessment Factors to Derive a DNEL, Brussels, Belgium.

ECHA, European Chemicals Agency. (2012). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. Version 2.1, November 2012, Helsinki, Finland. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment.

Hall, A.P., Elcombe, C.R., Foster, J.R., Harada, T., Kaufmann, W., Knippel, A., Küttler, K., Malarkey, D.E., Maronpot, R.R., Nishikawa, A., Nolte, T., Schulte, A., Strauss, V., York, M.J. (2012). Liver hypertrophy: a review of adaptive (adverse and

non-adverse) changes - conclusions from the 3rd International ESTP Expert Workshop. Toxicologic Pathology 40, 971-994.

Hartwig, A. (2020). 1,1,2,2-Tetrachlorethan, Nachtrag 2020, Entwurfsfassung. Gesundheitsschädliche Arbeitsstoffe, Toxikologisch-arbeitsmedizinische Begründungen von MAK-Werten, Loseblattsammlung. DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag Weinheim.

Hartwig, A., MAK Commission. (2018). Benzoic acid and alkali benzoates. The MAK Collection for Occupational Health and Safety 2017, Vol. 3, No. 4, DFG Deutsche Forschungsgemeinschaft, WILEY-VCH Verlag GmbH & Co. KGaA.

Nielsen, G.D., Wolkoff, P. (2017). Evaluation of airborne sensory irritants for setting exposure limits or guidelines: A systematic approach. Regulatory Toxicology and Pharmacology 90, 308-317.

NTP, National Toxicology Program. (2004). Technical Report on the Toxicity Studies of 1,1,2,2-Tetrachloroethane Administered in Microcapsules in Feed to F344/N Rats and B6C3F1 Mice. Toxicity Report Series 49, U.S. Department of Health and Human Services; Public Health Service. https://ntp.niehs.nih.gov/ntp/htdocs/st_rpts/tox049.pdf.

Pacharra, M., Kleinbeck, S., Schäper, M., Juran, S.A., Hey, K., Blaszkewicz, M., Lehmann, M.L., Golka, K., van Thriel, C. (2016). Interindividual differences in chemosensory perception: Toward a better understanding of perceptual ratings during chemical exposures. Journal of Toxicology and Environmental Health. Part A 79, 1026-1040.

Schneider, K., Hassauer, M., Oltmanns, J., Schuhmacher-Wolz, U., Elmshäuser, E., Mosbach-Schulz, O. (2004). Wahrscheinlichkeitsrechnung als Hilfsmittel zu Wirkungsabschätzung bei Arbeitnehmern. Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Fb 1012, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund/Berlin/Dresden, Wirtschaftsverlag NW Bremerhaven.

Schneider, K., Schuhmacher-Wolz, U., Hassauer, M., Darschnik, S., Elmshäuser, E., Mosbach-Schulz, O. (2006). A probabilistic effect assessment model for hazardous substances at the workplace. Regulatory Toxicology and Pharmacology 44, 172-181.

Schröder, K., Escher, S., Licht, O., Simetska, N., Mangelsdorf, I. (2015). Time Extrapolation Factors for Local Effects in Inhalation Studies. Project number [40191]. Draft March 2015, Fraunhofer Institute for Toxicology and Experimental Medicine ITEM, Chemical Risk Assessment, Databases and Expert Systems. On behalf of Umweltbundesamt, Germany.

Shusterman, D., Balmes, J. (1997). Measurement of nasal irritant sensitivity to pulsed carbon dioxide: a pilot study. Archives of Environmental Health 52, 334-340.

Sucker, K., Hoffmeyer, F., Monse, C., Jettkant, B., Berresheim, H., Rosenkranz, N., Raulf, M., Bunger, J., Bruning, T. (2019). Ethyl acrylate: influence of sex or atopy on perceptual ratings and eye blink frequency. Archives of Toxicology 93, 2913-2926.

WHO, World Health Organization. (2014). Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. IPCS Harmonization Project Document 11, WHO, Geneva, Switzerland. http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf.

List of Tables

List of Figures

Annex 1: R script for fitting an empirical distribution with the package rriskDistribution

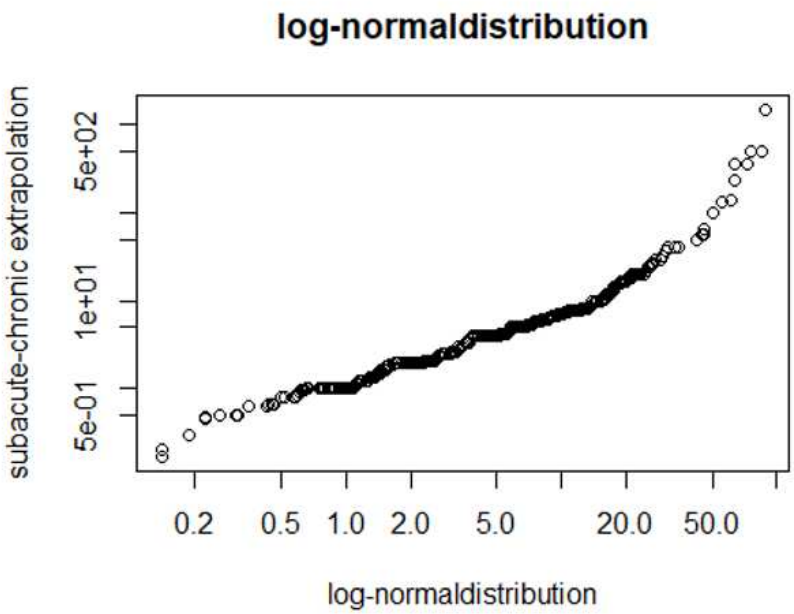
Exemplary shown for the subacute-chronic time extrapolation.

```
library(rriskDistributions)

# data loading
# for the purpose of this script, the values of the empirical distribution
are pasted here as a (sorted) vector

sa_c_ntp <- c(0.17, 0.208666666666667, 0.3, 0.46, 0.4890625, 0.5, 0.5, 0.5
, 0.625, 0.625, 0.666666666666667, 0.666666666666667, 0.666666666666667, 0
.8, 0.8, 0.814285714285714, 0.814285714285714, 0.833333333333333, 0.944444
444444444, 0.96875, 0.978125, 0.992, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
,1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1.04166666666667, 1.06666
666666667, 1.06666666666667, 1.1144578313253, 1.2, 1.2, 1.2175, 1.2175,1.2
4, 1.24, 1.25, 1.25, 1.25, 1.25, 1.33333333333333, 1.33333333333333, 1.333
33333333333, 1.33333333333333, 1.33333333333333, 1.42, 1.5, 1.5, 1.5,1.562
5, 1.625, 1.66666666666667, 1.66666666666667, 1.66666666666667, 1.66666666
666667, 1.66666666666667, 1.66666666666667, 1.77777777777778, 1.7857142857
1429, 1.9, 1.92307692307692, 1.92307692307692, 1.95625, 1.96875, 1.9808306
7092652, 1.98083067092652, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2.01612903225806, 2.
03225806451613, 2.08333333333333, 2.08333333333333, 2.08333333333333, 2.08
333333333333, 2.08333333333333, 2.08333333333333, 2.08333333333333, 2.0833
3333333333, 2.10666666666667, 2.1875, 2.34375, 2.42857142857143, 2.435, 2.
44444444444444, 2.48, 2.48, 2.48, 2.48, 2.5, 2.5, 2.5, 2.5, 2.5, 2.5, 2.5,
2.5, 2.5, 2.5, 2.5, 2.5, 2.5, 2.5, 2.5, 2.5, 2.50666666666667, 2.659685863
87435, 2.66666666666667, 2.66666666666667, 2.66666666666667, 2.66666666666
667, 2.66666666666667, 2.77777777777778, 3, 3, 3, 3, 3, 3, 3.125, 3.166666
66666667, 3.25, 3.3, 3.33333333333333, 3.3433734939759, 3.4, 3.73333333333
333, 3.875, 3.97435897435897, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4.03225806451613, 4.
03225806451613, 4.03225806451613, 4.03571428571429, 4.16666666666667, 4.16
666666666667, 4.16666666666667, 4.16666666666667, 4.16666666666667, 4.1666
6666666667, 4.16666666666667, 4.16666666666667, 4.16666666666667, 4.166666
66666667, 4.41666666666667, 4.59016393442623, 4.76190476190476, 4.76190476
190476, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5.3333333
3333333, 5.33333333333333, 5.33333333333333, 5.33333333333333, 5.333333333
33333, 5.33333333333333, 5.5, 5.71428571428571, 5.8, 6, 6, 6, 6, 6, 6.0096
1538461539, 6.23809523809524, 6.25, 6.25, 6.25, 6.25, 6.25, 6.25, 6.266666
66666667, 6.26666666666667, 6.5, 6.5, 6.66666666666667, 6.66666666666667,
6.66666666666667, 6.66666666666667, 7, 7, 7, 7, 7.25, 7.31331793687452, 7.
33333333333333, 7.5, 7.5, 7.6, 7.6, 7.875, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8
, 8, 8, 8, 8, 8, 8.01282051282051, 8.06451612903226, 8.33333333333333, 8.3
```

```
3333333333333, 8.33333333333333, 8.33333333333333, 8.33333333333333, 8.333
33333333333, 8.33333333333333, 8.57142857142857, 10, 10, 10, 10, 10, 10, 1
0, 10, 10, 10.2264150943396, 10.28, 10.3617021276596, 10.8520179372197, 11
, 12, 12, 12, 12, 12.4468085106383, 12.5, 13.2978723404255, 13.33333333333
33, 14, 14.5, 15, 15.8333333333333, 16, 16, 16, 16.6666666666667, 16.66666
66666667,16.6666666666667, 16.6666666666667, 16.6666666666667, 17.85714285
71429, 18, 18.75, 20, 20, 20, 20, 20, 20, 20, 20, 21.8571428571429, 24, 24
, 25.9493670886076, 26.6666666666667, 26.6666666666667, 29.4117647058824,
29.4117647058824, 33, 37.5, 37.5, 40, 40, 40.3225806451613, 50, 56.25, 56.
25, 66.6666666666667, 100, 135, 142.857142857143, 240, 360, 360, 500, 500,
500, 1500)



summary(sa_c_ntp)

##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    0.170    2.000    4.000   17.079    7.906 1500.000
```

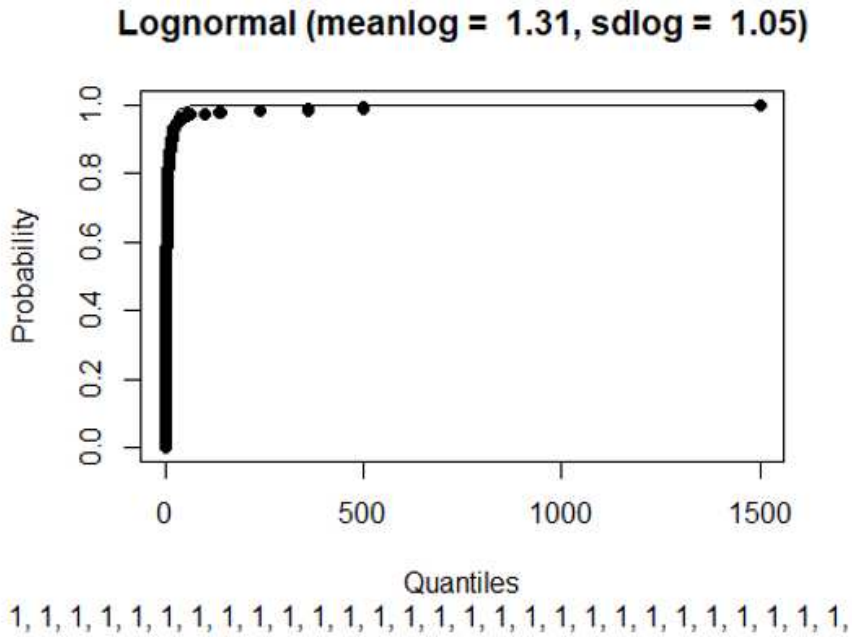
Check for Lognormality

```
qqplot(x = rlnorm(n = length(sa_c_ntp), meanlog = mean(log(sa_c_ntp)), sdl
og = sd(log(sa_c_ntp))), y = sa_c_ntp, main = "log-normaldistribution",  y
lab = "subacute-chronic extrapolation", xlab = "log-normaldistribution", l
og = "xy")
```



fitting to lognormal distribution with rriskDistributions

```
# generate vector of probabilities
ptheo <- ((1:length(sa_c_ntp))-0.5)/length(sa_c_ntp)


# fit to lognorm distribution
rrisk_fit <- rriskDistributions::get.lnorm.par(p = ptheo, q = sort(sa_c_nt
p))

## The fitting procedure 'L-BFGS-B' was successful!

## $par
## [1] 1.309572 1.050037
##
## $value
## [1] 1.571309e-06
##
## $counts
## function gradient
##        9        9
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```



1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,

```r
# get mu
sa_c_ntp_mu <- rrisk_fit[1]
sa_c_ntp_mu

##   meanlog
## 1.309572

# get sigma
sa_c_ntp_sigma <- rrisk_fit[2]
```

Annex 2: Probabilistic evaluation of example substances

Simulation reports of the EFSA MonteCarlo Tool (https://shiny-efsa.openanalytics.eu/) for 5% incidence.

Example 1: 1,1,2,2-Tetrachloroethane

Example 2: Benzoic acid

Monte Carlo Report

Simulation

July 22, 2020

Contents

1 MODEL

1.1 Variables

Table 1: Table continues below

	Abbr	Label
1	oel	OEL
11	td5	intraspecies-TD 5%
2	logGSD	intraspecies-TK log GSD
3	z95	z-score-95
4	bmd	BMD
5	time_sc_c	time extrapolation subchronic to chronic
6	inter	interspecies
7	pod_adjust	PoD adjustment

	Description	Unit	Distribution
1	Simulation of OEL distribution	mg/m3	
11	Distribution of uncertainty factors for the toxicodynamic part of intraspecies variability. Distribution according to a 5% incidence level in the population		a+LOGNORMAL
2	Distribution of log10 GSD values derived from toxicokinetic studies to derive the toxicokinetic part of intraspecies variability		a+LOGNORMAL
3	z-score corresponding to 95% coverage		CONSTANT(a)
4	Distribution of BMD for liver weight changes by 1,1,2,2-Tetrachloroethane. Derived from the BMD_L and BMD_U output of the dose response modelling, assuming a lognormal distribution.		a+LOGNORMAL
5	Distribution for the time extrapolation subchronic to chronic		a+LOGNORMAL
6	Distribution for interspecies extrapolation		a+LOGNORMAL
7	(absorption animal, oral / absorption human, inhalation) * (7 exposure days per week animals / 5 exposure days per week workers) * (allometric scaling) * (70 kg bodyweight /10 m³ breathing volume per workday) = (0.95/0.60) * (7/5) * (1/4) * (70/10) = 3.879		CONSTANT(a)

1.2 Model Equation

$$oel = \frac{bmd * pod\_adjust}{(time\_sc\_c * td5 * 10^{(logGSD * z95)} * inter)}$$

2 SIMULATION

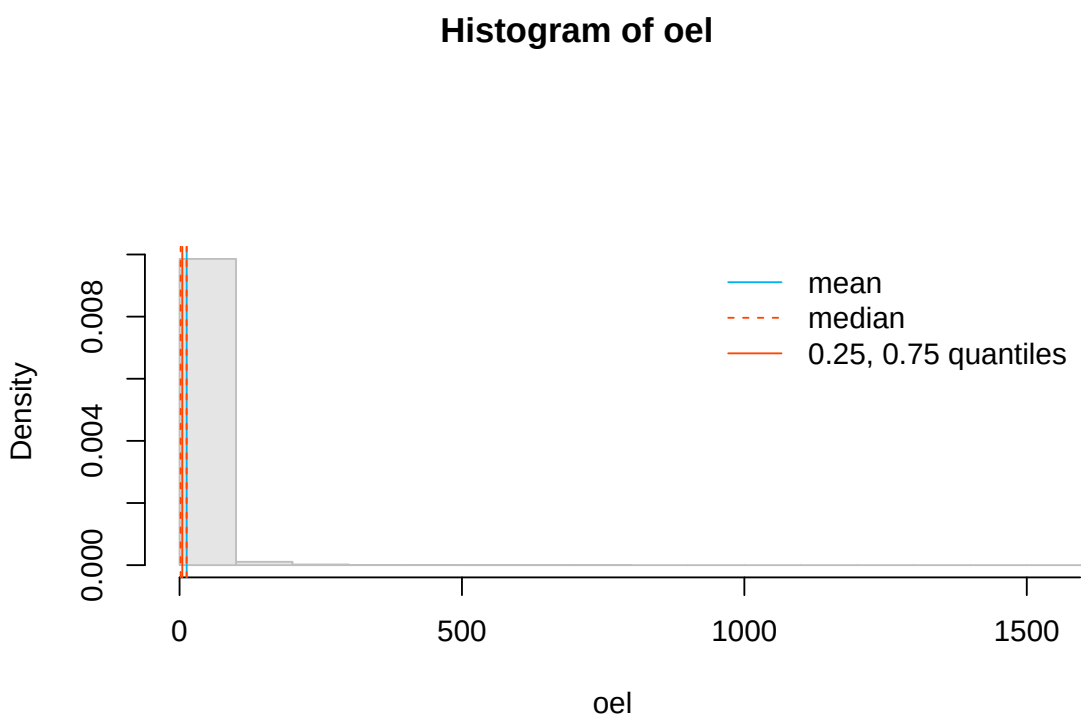
Table 3: Table continues below

variable	mean	sd	0.5%	1%	5%	10%	25%
td5	2.072	0.7927	0.7471	0.8168	1.052	1.206	1.51
logGSD	0.1756	0.1185	0.0301	0.03506	0.05301	0.06649	0.09655
z95	1.645	0	1.645	1.645	1.645	1.645	1.645
bmd	14.15	4.73	5.841	6.332	7.878	8.847	10.78
time_sc_c	4.638	6.192	0.2198	0.2856	0.5575	0.8026	1.447
inter	1.344	1.154	0.1446	0.1759	0.297	0.388	0.6126
pod_adjust	3.879	0	3.879	3.879	3.879	3.879	3.879
oel	12.63	29.57	0.1113	0.1673	0.4693	0.7952	1.901

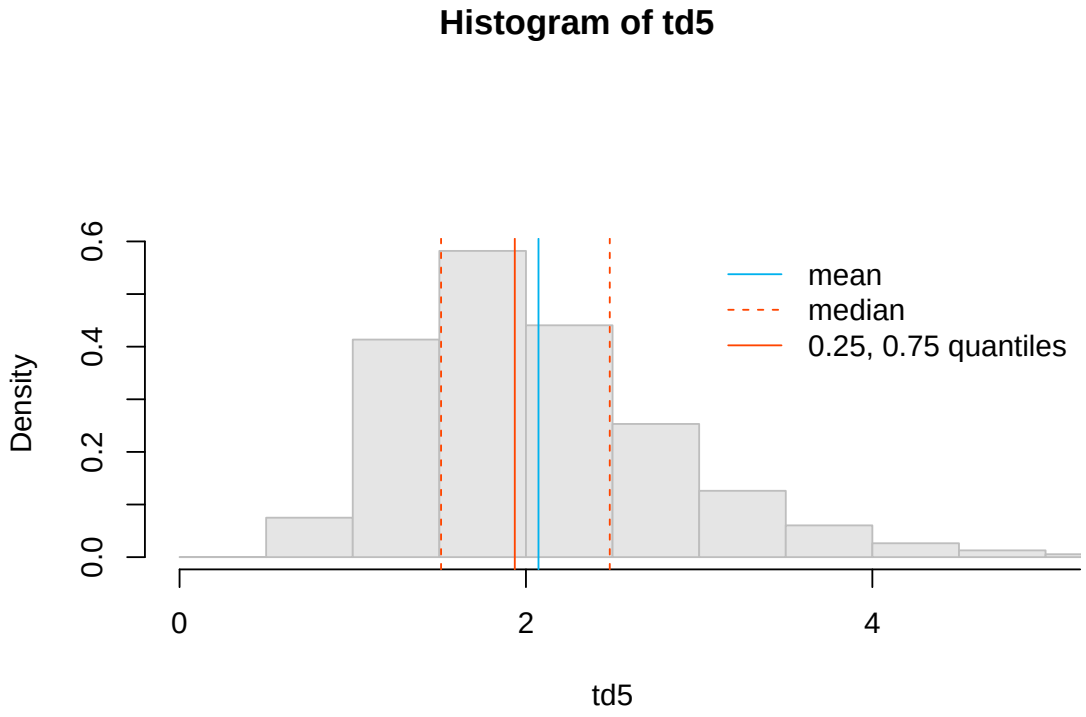
33%	50%	66%	75%	90%	95%	99%	99.5%	Na's
1.644	1.935	2.255	2.484	3.109	3.554	4.59	4.995	0
0.1113	0.1456	0.1872	0.2196	0.3199	0.3999	0.5984	0.7028	0
1.645	1.645	1.645	1.645	1.645	1.645	1.645	1.645	0
11.64	13.43	15.33	16.69	20.33	22.9	28.59	31.06	0
1.818	2.817	4.265	5.521	10.13	14.41	28.69	36.88	0
0.7319	1.015	1.386	1.686	2.655	3.48	5.798	6.921	0
3.879	3.879	3.879	3.879	3.879	3.879	3.879	3.879	0
2.634	4.88	8.639	12.49	28.66	47	118.5	168.8	0

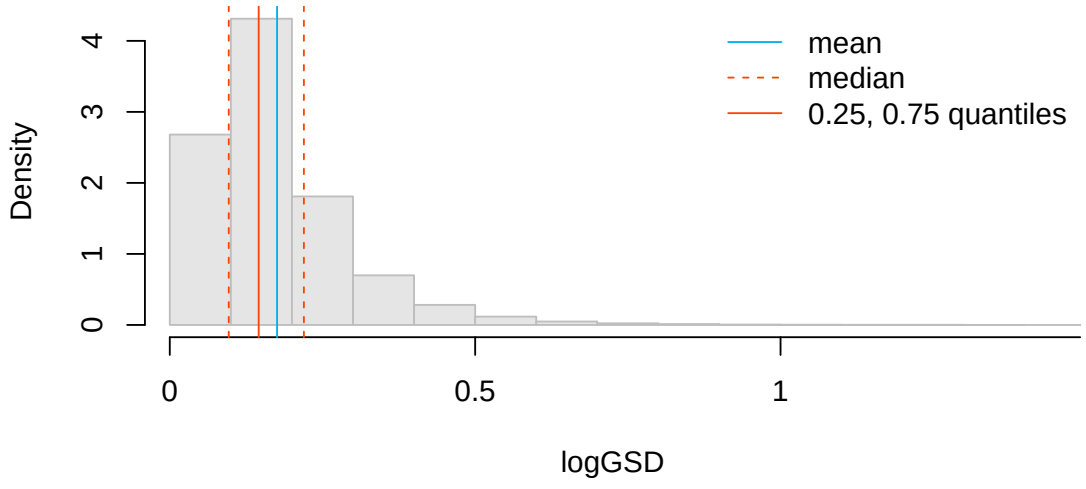
2.1 Output Histogram

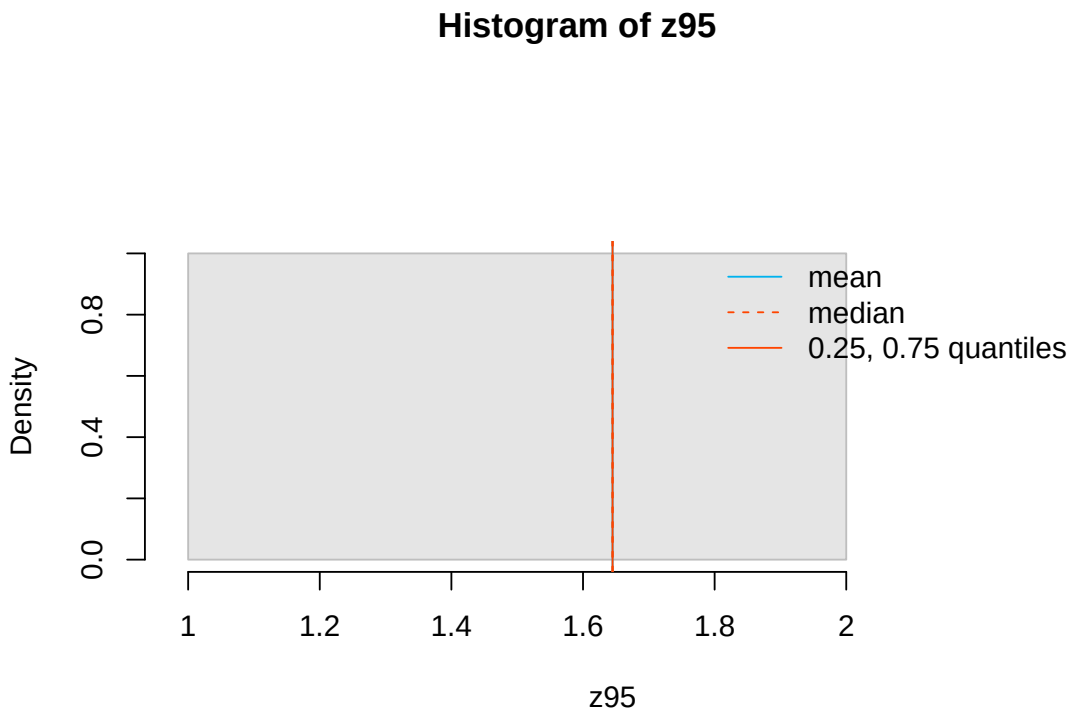
2.1.1 Output Variable

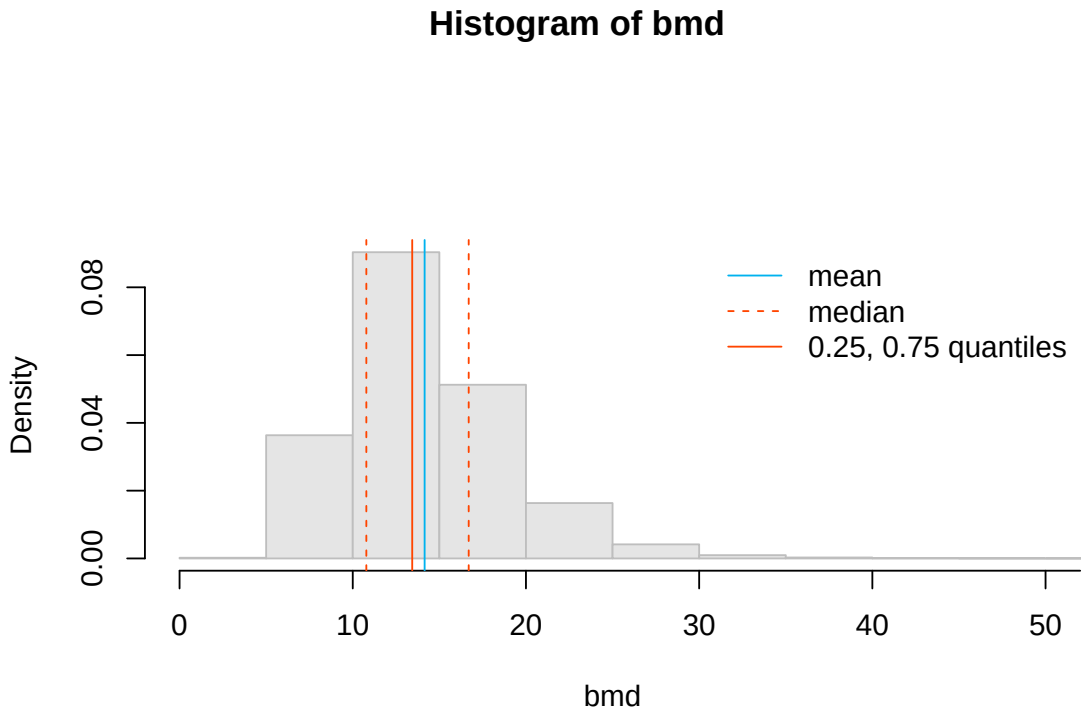


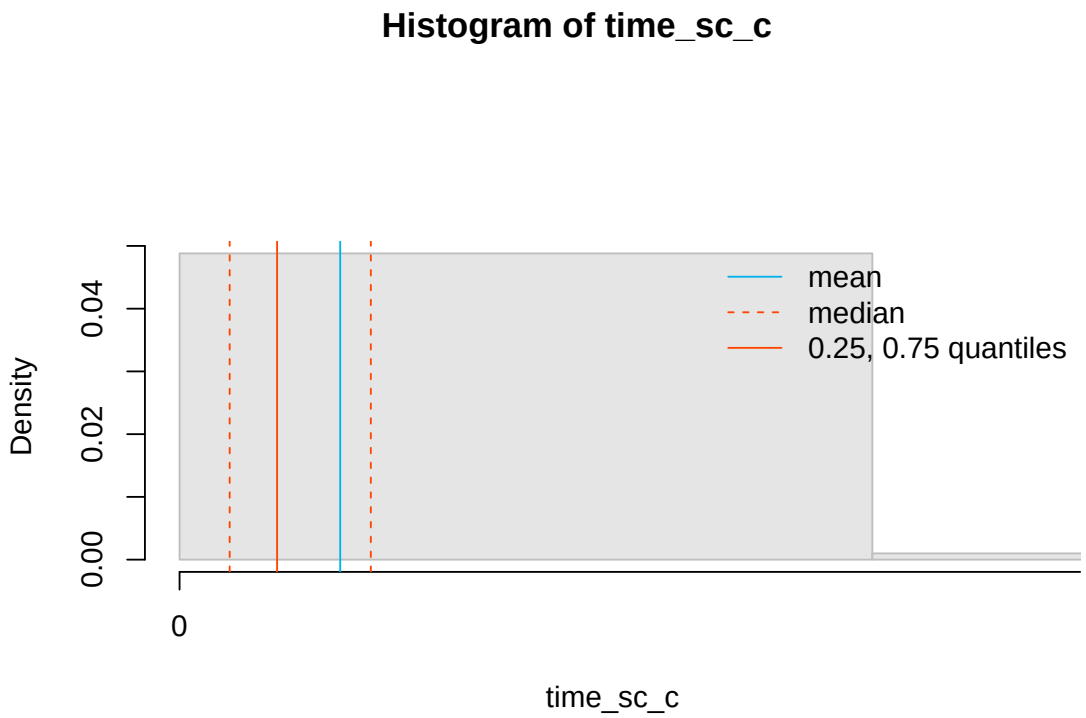
2.1.2 Input Variables

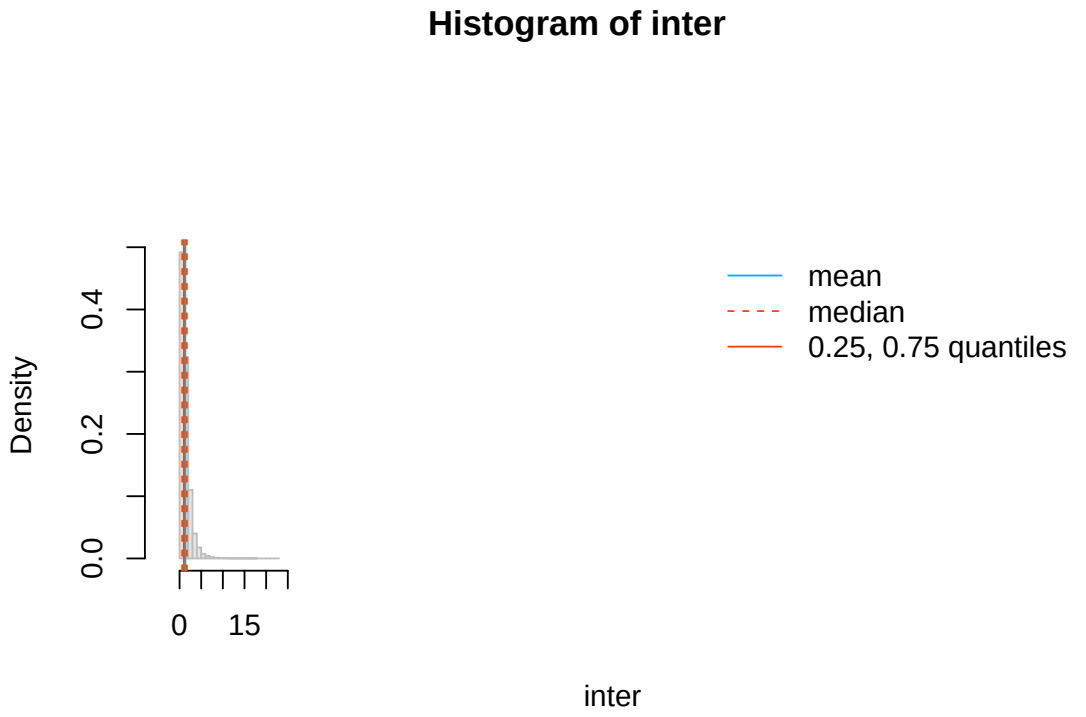


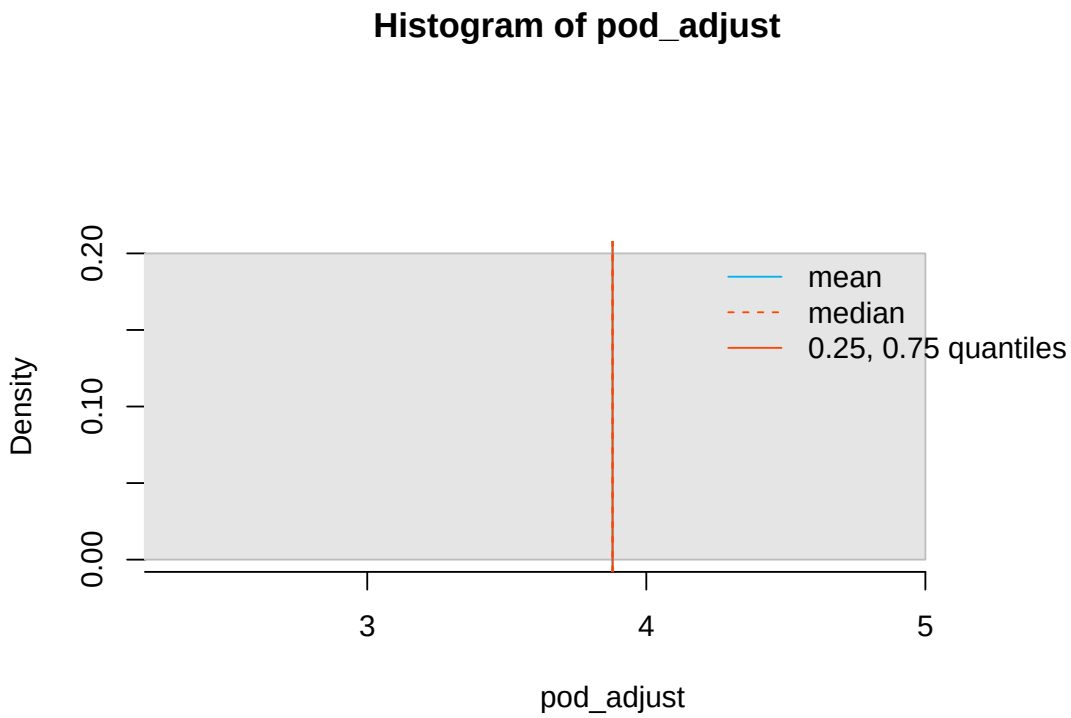



Histogram of z95
Density
0.0
0.4
0.8
1
1.2
1.4
1.6
1.8
2
z95
mean
median
0.25, 0.75 quantiles


Histogram of bmd
Density
0.00
0.04
0.08
0
10
20
30
40
50
bmd
mean
median
0.25, 0.75 quantiles


Histogram of time_sc_c
Density
0.00
0.02
0.04
0
mean
median
0.25, 0.75 quantiles
time_sc_c

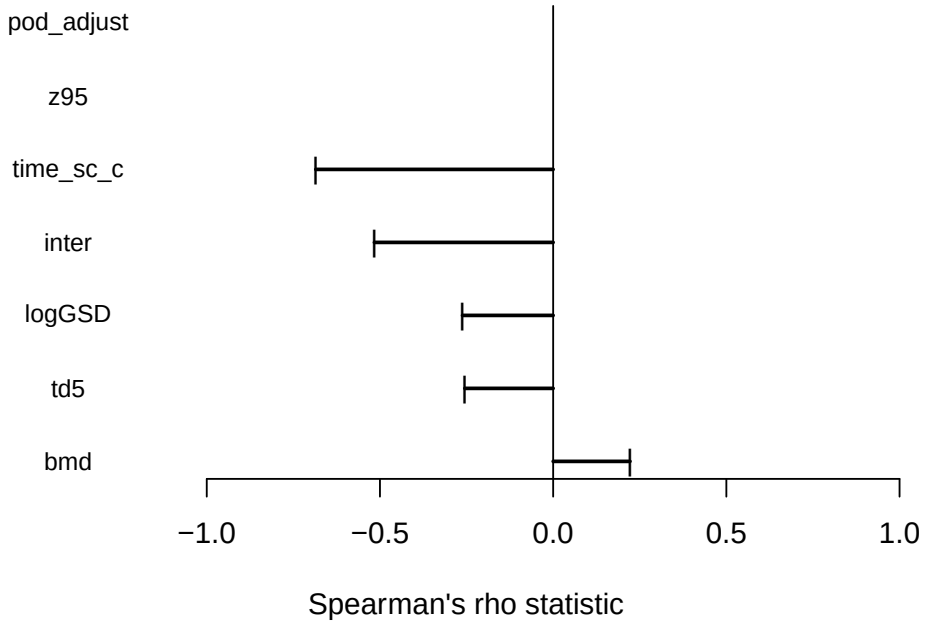

Histogram of inter
Density
0.0
0.2
0.4
0
15
mean
median
0.25, 0.75 quantiles
inter



2.2 Sensitivity

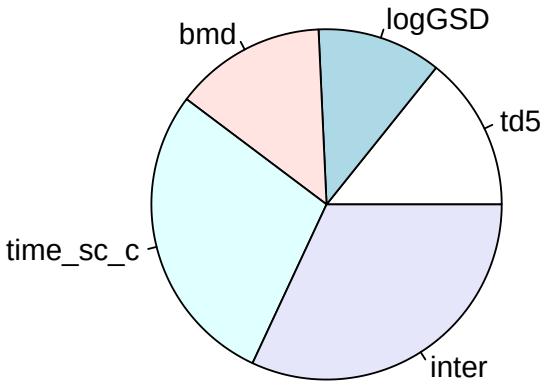
2.2.1 Tornado Chart

```
## Warning in cor(out[, y, ], inp[, ifelse(nunci == 1, 1, y), ], method = method, :
## the standard deviation is zero

## Warning in cor(out[, y, ], inp[, ifelse(nunci == 1, 1, y), ], method = method, :
## the standard deviation is zero
```



2.2.2 Pie Chart

```
## Warning in sensitivityToVarianceRatios(sensitivityObj): The input variable(s)
## z95, pod_adjust are collinear and have been ommitted.
```


bmd
logGSD
td5
time_sc_c
inter

Monte Carlo Report

Simulation

July 22, 2020

Contents

1 MODEL

1.1 Variables

Table 1: Table continues below

	Abbr	Label
1	oel	OEL
11	td5	intraspecies-TD 5%
2	logGSD	intraspecies-TK log GSD
3	z95	z-score-95
4	bmd	BMD
5	time_sa_c	time extrapolation subacute to chronic
6	inter	interspecies
7	pod_adjust	PoD adjustment

	Description	Unit	Distribution
1	Simulation of OEL distribution	mg/m3	
11	Distribution of uncertainty factors for the toxicodynamic part of intraspecies variability. Distribution according to a 5% incidence level in the population		a+LOGNORMAL
2	Distribution of log10 GSD values derived from toxicokinetic studies to derive the toxicokinetic part of intraspecies variability		a+LOGNORMAL
3	Z-score corresponding to 95% coverage		CONSTANT(a)
4	Distribution of BMD for lung irritation by benzoic acid. Derived from the BMD_L and BMD_U output of the dose response modelling, assuming a lognormal distribution.		a+LOGNORMAL
5	Distribution for the time extrapolation subacute to chronic		a+LOGNORMAL
6	Distribution for interspecies extrapolation		a+LOGNORMAL
7	(5 exposure days per week animals / 5 exposure days per week workers) * (6 h exposure per day animals / 8 h exposure per day workers) * (daily breathing volume in rest / daily breathing volume under light physical activity) = (5/5) * (6/8) * (6.7/10) = 0.5025		CONSTANT(a)

1.2 Model Equation

$$oel = \frac{bmd * pod\_adjust}{(time\_sa\_c * td5 * 10^{(logGSD * z95)} * inter)}$$

2 SIMULATION

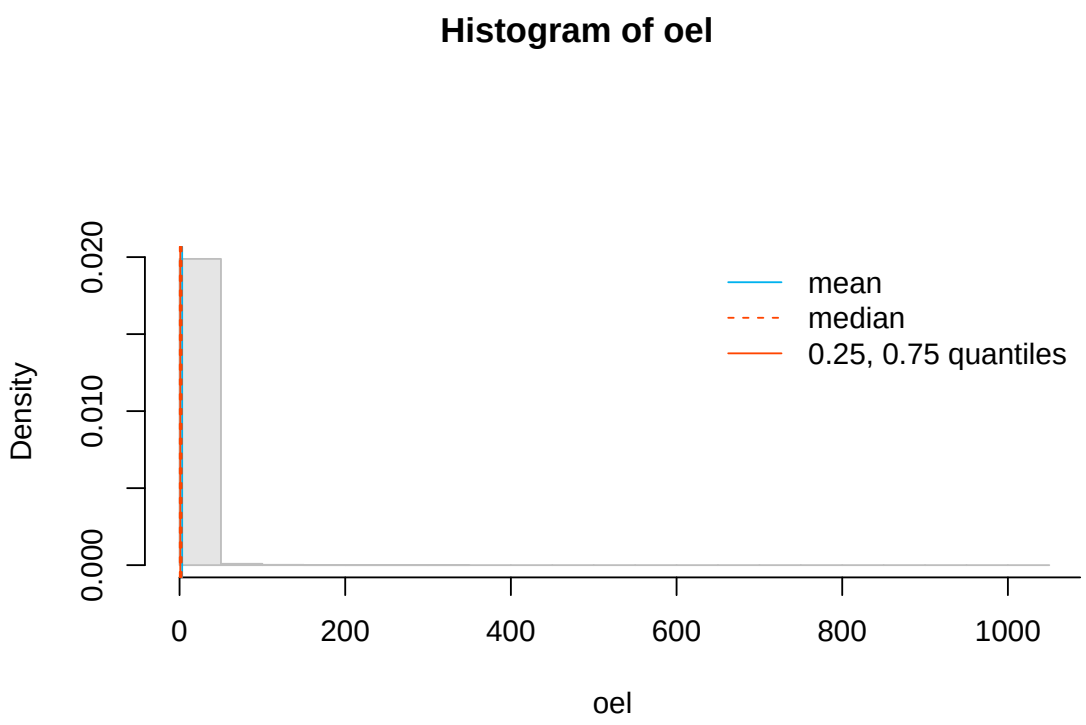
Table 3: Table continues below

variable	mean	sd	0.5%	1%	5%	10%	25%
td5	2.072	0.7927	0.7471	0.8168	1.052	1.206	1.51
logGSD	0.1756	0.1185	0.0301	0.03506	0.05301	0.06649	0.09655
z95	1.645	0	1.645	1.645	1.645	1.645	1.645
bmd	33.67	33.08	3.017	3.692	6.376	8.521	13.99
time_sa_c	6.464	9.564	0.2466	0.3256	0.6618	0.9741	1.82
inter	1.344	1.154	0.1446	0.1759	0.297	0.388	0.6126
pod_adjust	0.5025	0	0.5025	0.5025	0.5025	0.5025	0.5025
oel	3.153	11.24	0.01126	0.01839	0.05795	0.1062	0.2899

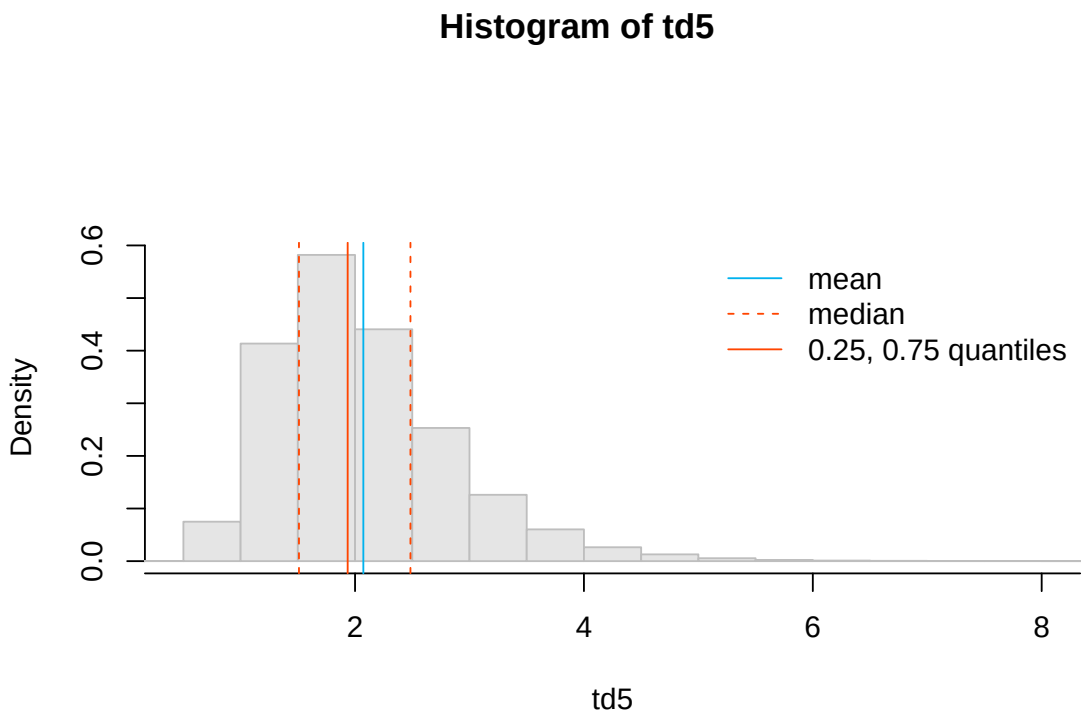
33%	50%	66%	75%	90%	95%	99%	99.5%	Na's
1.644	1.935	2.255	2.484	3.109	3.554	4.59	4.995	0
0.1113	0.1456	0.1872	0.2196	0.3199	0.3999	0.5984	0.7028	0
1.645	1.645	1.645	1.645	1.645	1.645	1.645	1.645	0
16.91	24.23	33.73	41.72	68.31	91.96	160.3	197.2	0
2.319	3.689	5.727	7.532	14.33	20.84	43.25	56.44	0
0.7319	1.015	1.386	1.686	2.655	3.48	5.798	6.921	0
0.5025	0.5025	0.5025	0.5025	0.5025	0.5025	0.5025	0.5025	0
0.4252	0.863	1.692	2.576	6.834	12.1	36.02	55.5	0

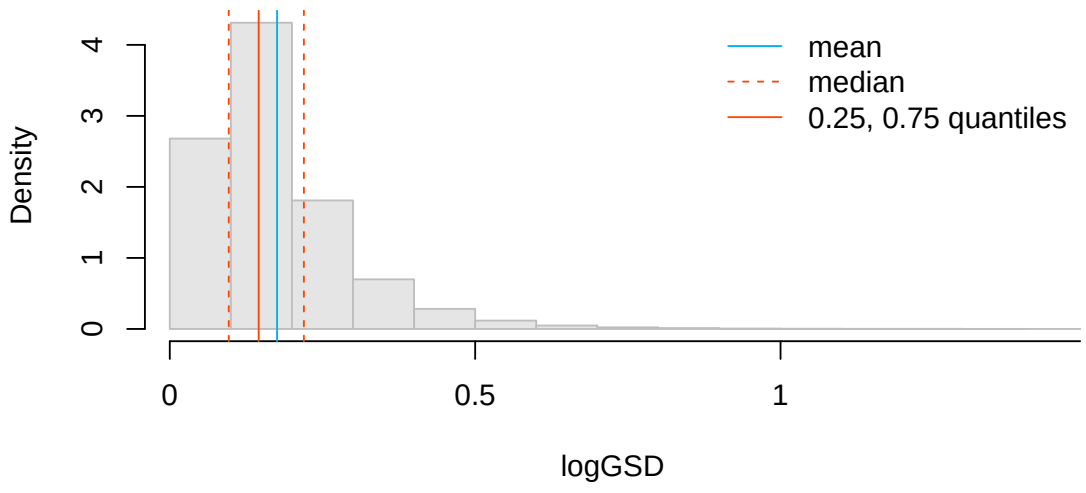
2.1 Output Histogram

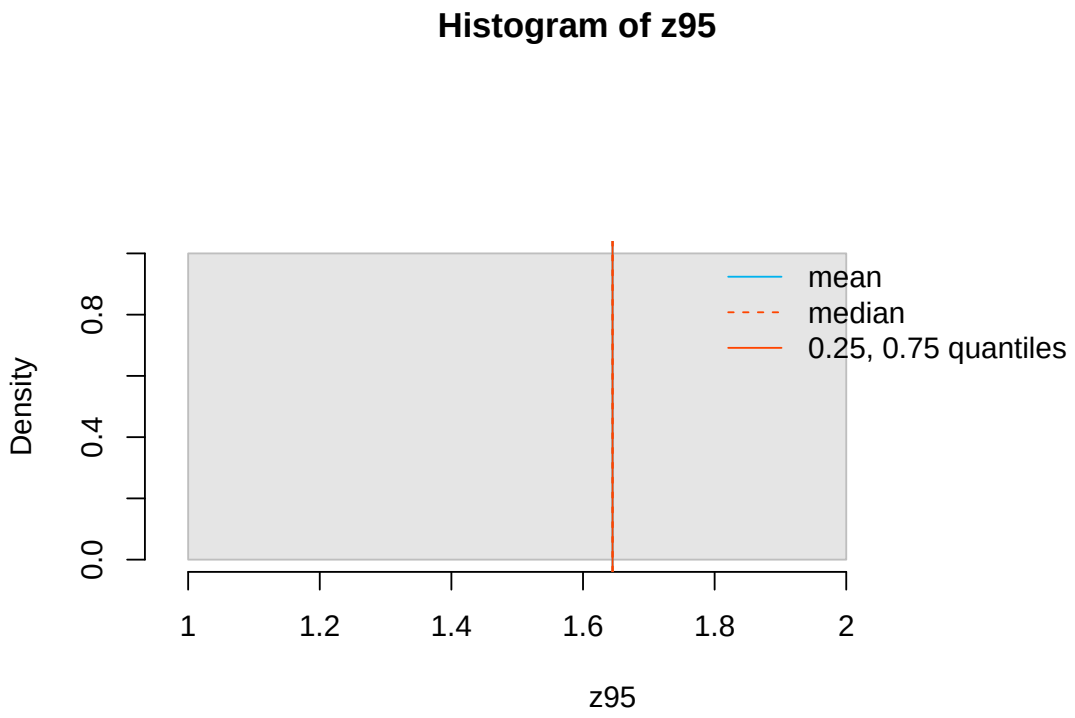
2.1.1 Output Variable

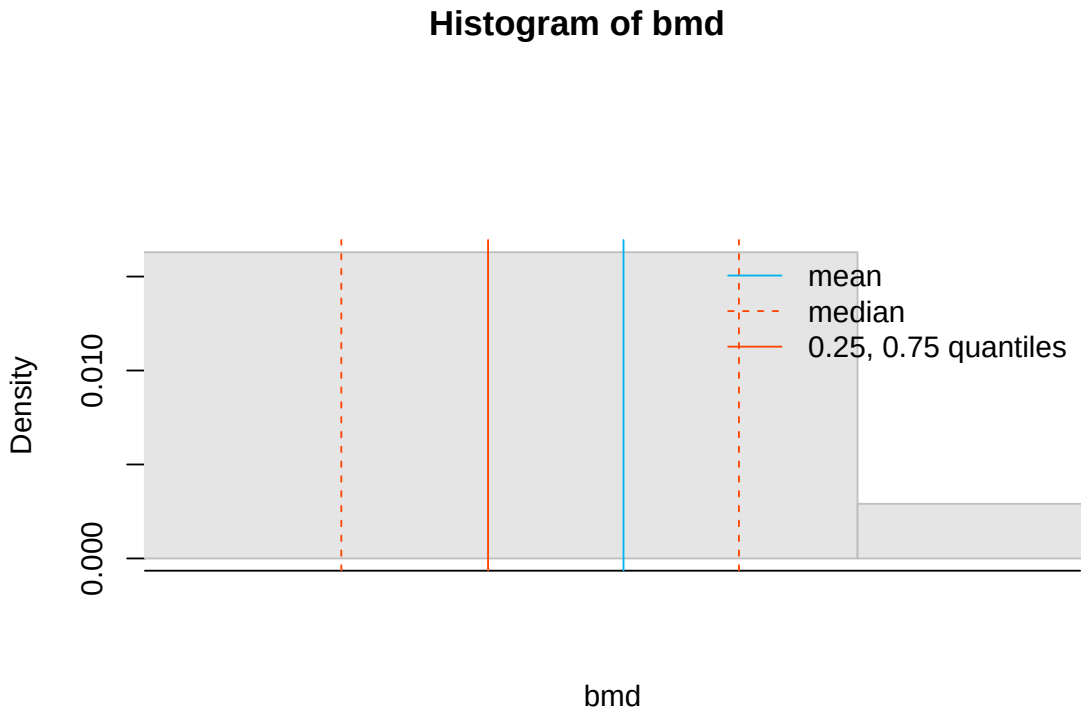


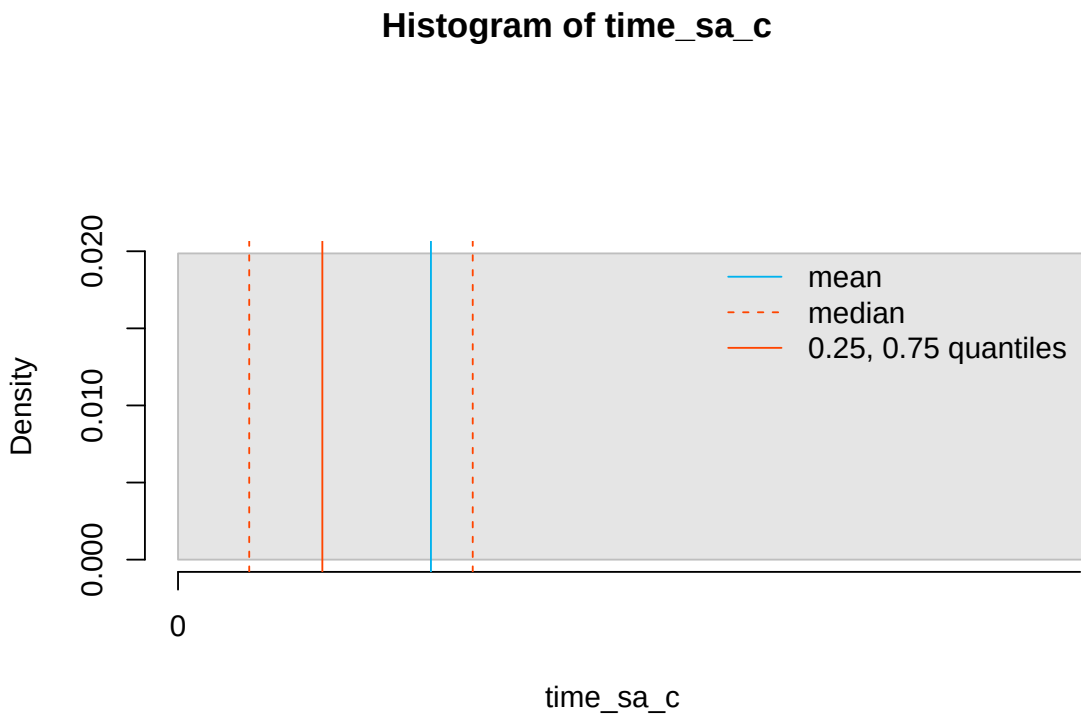
2.1.2 Input Variables

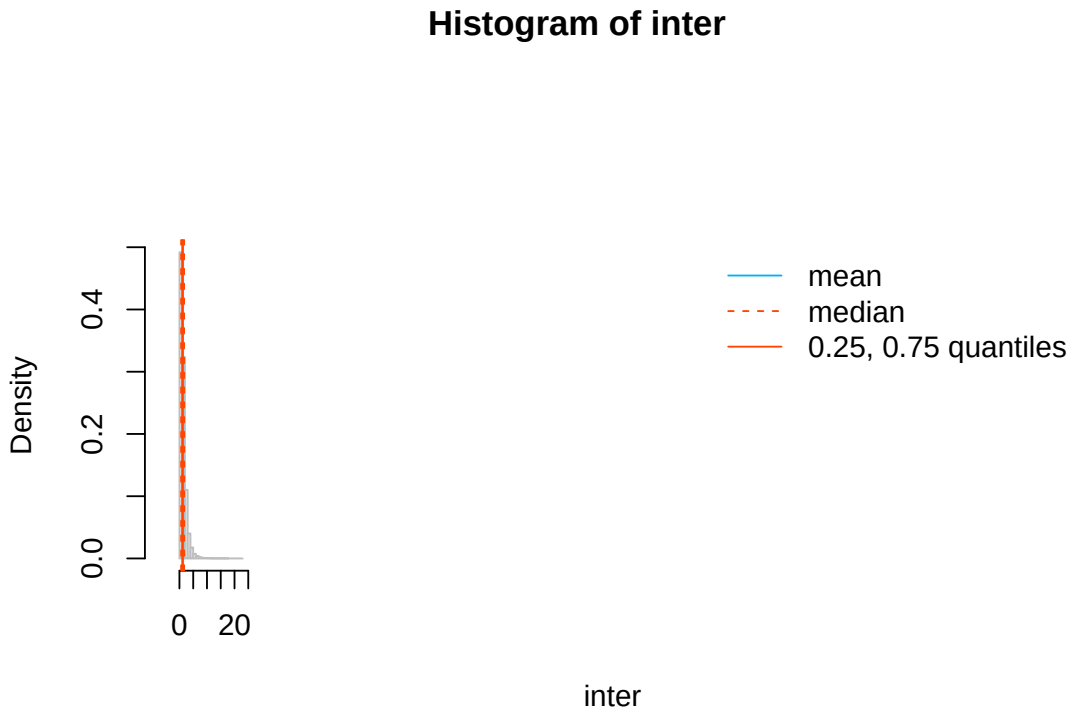


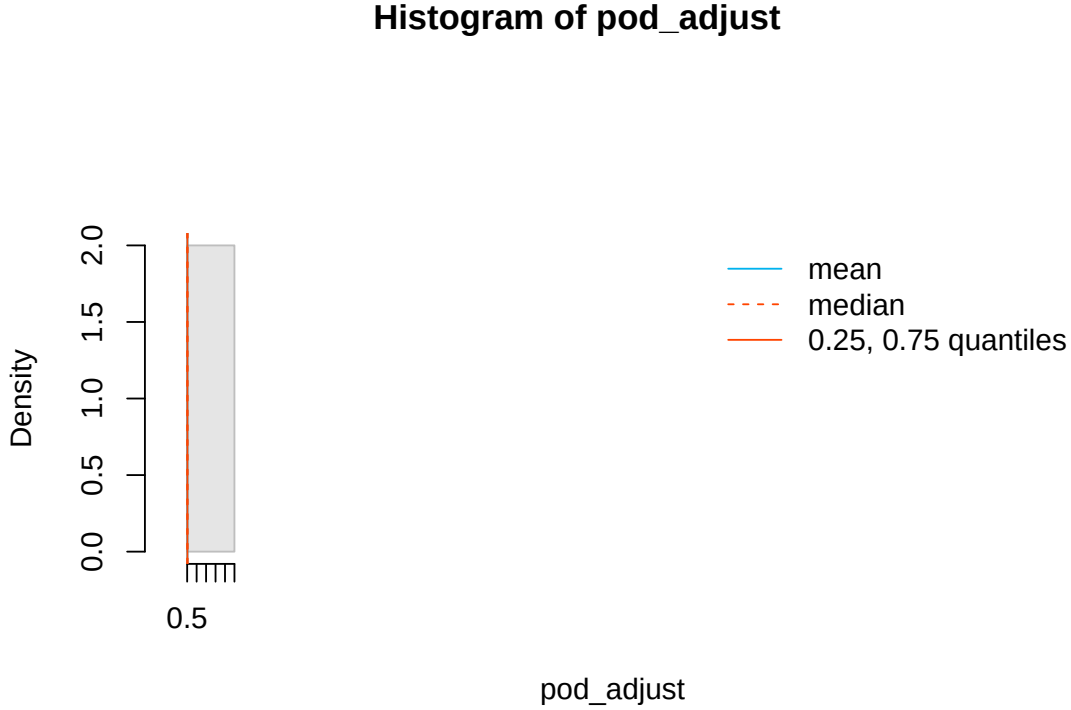



Histogram of z95
Density
0.0
0.4
0.8
1
1.2
1.4
1.6
1.8
2
z95
mean
median
0.25, 0.75 quantiles


Histogram of bmd
Density
0.000
0.010
bmd
mean
median
0.25, 0.75 quantiles


Histogram of time_sa_c
Density
0.000
0.010
0.020
0
mean
median
0.25, 0.75 quantiles
time_sa_c

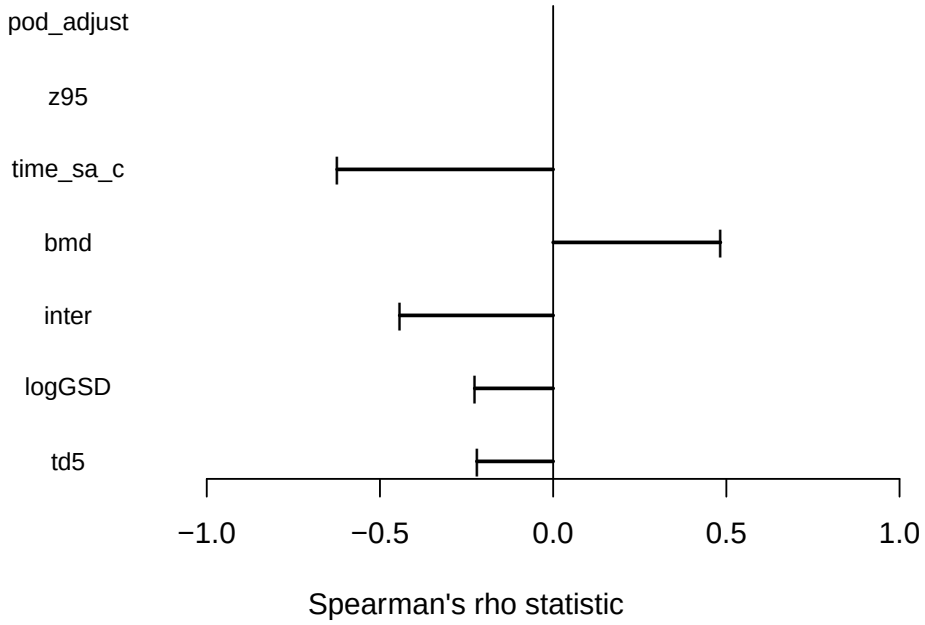

Histogram of inter
Density
0.0
0.2
0.4
0
20
mean
median
0.25, 0.75 quantiles
inter



2.2 Sensitivity

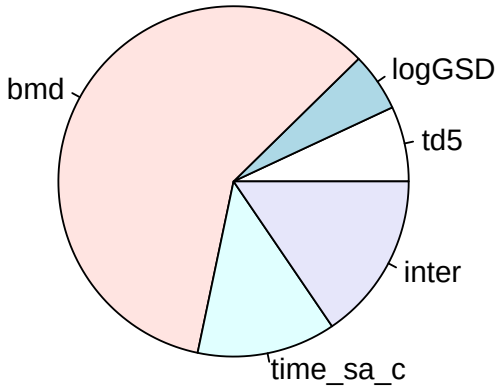
2.2.1 Tornado Chart

```
## Warning in cor(out[, y, ], inp[, ifelse(nunci == 1, 1, y), ], method = method, :
## the standard deviation is zero

## Warning in cor(out[, y, ], inp[, ifelse(nunci == 1, 1, y), ], method = method, :
## the standard deviation is zero
```



2.2.2 Pie Chart

```
## Warning in sensitivityToVarianceRatios(sensitivityObj): The input variable(s)
## z95, pod_adjust are collinear and have been ommitted.
```


bmd
logGSD
td5
inter
time_sa_c