



Probabilistic Points of Departure and Reference Doses for Characterizing Human Noncancer and Developmental/Reproductive Effects for 10,145 Chemicals

Aurisano, Nicolò; Jolliet, Olivier; Chiu, Weihsueh A.; Judson, Richard; Jang, Suji; Unnikrishnan, Aswani; Kosnik, Marissa B.; Fantke, Peter

Published in:
Environmental Health Perspectives

Link to article, DOI:
[10.1289/EHP11524](https://doi.org/10.1289/EHP11524)

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Aurisano, N., Jolliet, O., Chiu, W. A., Judson, R., Jang, S., Unnikrishnan, A., Kosnik, M. B., & Fantke, P. (2023). Probabilistic Points of Departure and Reference Doses for Characterizing Human Noncancer and Developmental/Reproductive Effects for 10,145 Chemicals. *Environmental Health Perspectives*, 131(3), Article 37016. <https://doi.org/10.1289/EHP11524>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Probabilistic Points of Departure and Reference Doses for Characterizing Human Noncancer and Developmental/Reproductive Effects for 10,145 Chemicals

Nicolò Aurisano,¹ Olivier Jolliet,^{1,2} Weihsueh A. Chiu,³ Richard Judson,⁴ Suji Jang,³ Aswani Unnikrishnan,⁴ Marissa B. Kosnik,¹ and Peter Fantke¹

¹Quantitative Sustainability Assessment, Department of Environmental and Resource Engineering, Technical University of Denmark, Lyngby, Denmark

²Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA

³Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA

⁴National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA

BACKGROUND: Regulatory toxicity values used to assess and manage chemical risks rely on the determination of the point of departure (POD) for a critical effect, which results from a comprehensive and systematic assessment of available toxicity studies. However, regulatory assessments are only available for a small fraction of chemicals.

OBJECTIVES: Using *in vivo* experimental animal data from the U.S. Environmental Protection Agency's Toxicity Value Database, we developed a semiautomated approach to determine surrogate oral route PODs, and corresponding toxicity values where regulatory assessments are unavailable.

METHODS: We developed a curated data set restricted to effect levels, exposure routes, study designs, and species relevant for deriving toxicity values. Effect levels were adjusted to chronic human equivalent benchmark doses (BMD_h). We hypothesized that a quantile of the BMD_h distribution could serve as a surrogate POD and determined the appropriate quantile by calibration to regulatory PODs. Finally, we characterized uncertainties around the surrogate PODs from intra- and interstudy variability and derived probabilistic toxicity values using a standardized workflow.

RESULTS: The BMD_h distribution for each chemical was adequately fit by a lognormal distribution, and the 25th percentile best predicted the available regulatory PODs [$R^2 \geq 0.78$, residual standard error (RSE) $\leq 0.53 \log_{10}$ units]. We derived surrogate PODs for 10,145 chemicals from the curated data set, differentiating between general noncancer and reproductive/developmental effects, with typical uncertainties (at 95% confidence) of a factor of 10 and 12, respectively. From these PODs, probabilistic reference doses (1% incidence at 95% confidence), as well as human population effect doses (10% incidence), were derived.

DISCUSSION: In providing surrogate PODs calibrated to regulatory values and deriving corresponding toxicity values, we have substantially expanded the coverage of chemicals from 744 to 8,023 for general noncancer effects, and from 41 to 6,697 for reproductive/developmental effects. These results can be used across various risk assessment and risk management contexts, from hazardous site and life cycle impact assessments to chemical prioritization and substitution. <https://doi.org/10.1289/EHP11524>

Introduction

Chemical management and assessment frameworks, whether for site cleanup, life cycle impact assessment (LCIA), chemical alternatives assessment (CAA), or comparative risk screening, all aim to evaluate toxicological impacts on human health from chemical exposures.^{1,2} These frameworks rely on chemical-specific points of departure (PODs) for deriving the quantitative toxicity values necessary for such evaluations. The POD represents the point on the dose–response curve marking the beginning of a low-dose extrapolation for risk assessment³ and is derived from effect levels from *in vivo* studies, such as the lowest observed adverse effect level (LOAEL), the no observed adverse effect level (NOAEL), and the statistically derived benchmark dose lower confidence limit (BMDL).⁴ Moreover, these PODs are typically required to be based on regulatory assessments that review and synthesize the

available toxicity data, such as the U.S. Environmental Protection Agency's (EPA's) Integrated Science Assessments and Integrated Risk Information System (IRIS) toxicological reviews and Provisional Peer Reviewed Toxicity Values (PPRTV), among others. Yet, regulatory data sources are only available for a very limited share of the several tens of thousands of chemical substances commonly used worldwide,^{5–7} mainly because developing such regulatory assessments is highly data-, time-, and resource-intensive.⁸ Regulatory assessment being generally based on the most sensitive end points, the number of chemicals with developmental/reproductive regulatory PODs is even more restricted.

For chemical risk assessment purposes, the World Health Organization International Programme on Chemical Safety (WHO/IPCS) developed a unified framework for dose–response assessment able to derive probabilistic reference doses (RfDs) from PODs.^{9–12} This framework provides a consistent and transparent approach for both health-based risk assessment as well as comparative risk. Moreover, in the LCIA context its implementation was recommended for deriving human dose–response factors for noncancer end points,¹ using human population effect doses with an incidence response level $I = 10\%$. However, the WHO/IPCS framework has only been applied to $n = 608$ substances with regulatory data to calculate probabilistic RfDs¹² and to $n = 115$ organic chemicals to calculate human population effect doses ($I = 10\%$).¹

With the increasing availability of online experimental animal databases, it is possible to obtain *in vivo* toxicity data for tens of thousands of chemical substances. Examples of such large toxicity data sources include the U.S. EPA's Toxicity Value Database (ToxValDB)¹³ and the International Uniform Chemical Information Database (IUCLID; <https://iuclid6.echa.europa.eu/>) developed under the European Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH) regulation (EC 1907/2006). We propose that through the application of rigorous curation and statistical approaches, these data sources can be used to derive

Address correspondence to Peter Fantke. Telephone: 45 45254452. Email: pefan@dtu.dk

Supplemental Material is available online (<https://doi.org/10.1289/EHP11524>).

While not affecting the method and results presented in this paper, O.J. discloses his role as a member of the USEtox Center scientific advisory board and chair of the project on Global Guidance for Life Cycle Impact Assessment a project supported by the Life Cycle Initiative, hosted at UN-Environment Programme. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. All other authors declare they have nothing to disclose.

Received 6 May 2022; Revised 6 February 2023; Accepted 3 March 2023; Published 29 March 2023.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehpsubmissions@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

“surrogate” animal-based PODs in a quantitative high-throughput approach, systematically evaluating separate PODs for both reproductive/developmental effects and nonreproductive/developmental effects. Specifically, for substances for which regulatory PODs are not available, such experimental animal data could be alternatively used to estimate a POD that closely mimics one that would be selected in a regulatory assessment context (Figure S1).¹

However, such an approach needs to address numerous challenges presented by these databases.¹⁴ For example, a chemical with multiple studies reported can have multiple effect-level values (i.e., experimental values of toxicity from individual studies) associated with it. A repeat dose toxicity data set for a single chemical may include several effect-level types (e.g., NOAELs, LOAELs) covering different observed critical effects (e.g., body weight, reproduction) for various tested species (e.g., rats, dogs), with orders of magnitude in the variability of the reported effect-level values.^{3,12,15} Thus, systematic methods for data selection and harmonization for human toxicity information, similar to those proposed for physico-chemical properties¹⁶ and freshwater ecotoxicity information,¹⁷ need to be developed.^{18,19} Given that regulatory PODs are intended to be protective of all potential adverse effects, the estimated POD should be at the lower end of the distribution of available toxicity values,²⁰ following careful data curation where needed.²¹

Therefore, our approach to expand the coverage of chemicals for which toxicity values could be derived consisted of four specific objectives, namely:

- To create a consistent and curated data set of chronic dose–response toxicity data for multiple noncancer end points for oral exposure
- To develop a statistical approach to determine oral PODs by comparing curated toxicity data against available regulatory values
- To provide an extended set of oral PODs with quantified uncertainties for a wide range of chemicals, differentiating between reproductive/developmental and nonreproductive/developmental effects
- To determine probabilistic RfDs for health-based or comparative risk assessments and human population effect doses ($I = 10\%$) for LCIA, both calculated from the extended set of oral PODs using the WHO/IPCS framework

Throughout this paper, we separately consider reproductive/developmental effects and nonreproductive/developmental effects (the latter hereafter referred to as “general noncancer effects”) owing to an average factor of roughly 20 difference in severity to affect human lifetime loss,^{1,22} as well as the differences in applicable life stages and exposure durations. The surrogate PODs we develop along with their corresponding probabilistic RfDs and human population effect doses are suitable for implementation into various chemical management and exposure and impact assessment frameworks for application in high-throughput risk screening, LCIA, CAA for chemical substitution, and exposure and risk prioritization.^{1,23,24}

Methods

Figure 1 provides an overview of the overall workflow followed in this paper. First, we curated and selected experimental animal toxicity data and split them into two distinct data sets covering general noncancer effects and reproductive/developmental effects (Figure 1A). Second, we collected POD values from regulatory data sources (POD_{reg}) (Figure 1B) and compared these POD_{reg} with the curated dose–response toxicity data to identify a statistical approach for deriving surrogate oral PODs (Figure 1C). Third, we systematically applied this approach to determine a surrogate POD for each substance in the two curated data sets

(Figure 1D). We then characterized the uncertainty around each of the surrogate PODs that was due to intrastudy and interstudy variability through a bootstrapping approach (Figure 1E). Finally, using the surrogate PODs and their uncertainty, we derived both probabilistic RfDs and human population effect doses ($I = 10\%$) for use in health-based or comparative risk assessments and LCIA, respectively (Figure 1F). The following sections detail each of these main steps.

Description of the *in Vivo* Input Data Set

The *in vivo* data were collected in March 2021 from the U.S. EPA’s ToxValDB (version 9.1), an experimental toxicity database compiled from >40 publicly available sources.¹³ These include—among others—the Toxicity Reference Database (ToxRefDB; version 2.0),^{25,26} IRIS (<https://www.epa.gov/iris>), Office of Pesticide Programs (OPP; <https://www.epa.gov/pesticides>), PPRTV (<https://www.epa.gov/pprtv>), European Chemicals Agency’s eChem Portal (<https://www.echemportal.org/echemportal>), and European Food Safety Authority’s Chemical Hazards Database (<https://www.efsa.europa.eu/en/data/chemical-hazards-data>). The current version of ToxValDB is accessible through the EPA’s CompTox Chemicals Dashboard (<https://comptox.epa.gov/dashboard>).²⁷ The accessed database contained 427,506 records providing toxicity information on >30,000 chemicals.

Input Data Curation and Selection

We curated and selected the toxicity data from the ToxValDB with a semiautomated process based on a set of specific criteria derived from the WHO/IPCS recommendations in dose–response modeling (Figure 1A).^{10–12} The curation aimed first to harmonize the reported information to facilitate the data processing in our study; second, to filter out all records not relevant for our analysis (e.g., exposure route different from oral); and third, to make reported toxicity animal data directly comparable across different tested species and study types. We summarize below the steps of the curation and selection process with a few examples and actions taken (e.g., filtering, extrapolation), and Tables S1–S3 detail the process, including additional examples and further explanations of the choices made.

1. **Effect-level types:** we focused on the three effect-level types used for deriving PODs (i.e., NOAELs, LOAELs, and BMDLs) and excluded all the records referring to other effect-level types. The curation included, for example, grouping effect levels reported as no effect level (NEL) and no observed effect level (NOEL) to NOAEL, or lowest effect level (LEL) and lowest observed effect level (LOEL) to LOAEL. In addition, we disregarded all records with NELs (or LELs) as effect-level types in all cases in which another record from the same study and with effect-level types equal to NOAEL (or LOAEL) was already available.
2. **Exposure route:** we focused on oral exposure as the route of interest in the present study and thus excluded all records referring to other routes. During the curation, we grouped exposure routes reported as “food,” “gavage,” “diet, unspecified,” “oral via capsule,” “drinking water,” “stomach intubation,” “oral, intragastric,” “oral, gavage,” “feed,” “diet,” “drinking water,” and “liquid diet” to oral. In cases of missing information, we assigned an exposure route as oral for those with reported units equal to milligrams per kilogram per day or equivalent.
3. **Effect values and units:** we converted reported effect values into a consistent unit of milligrams per kilogram per day. We excluded all the records with missing effect values or unconvertible and unclear units (e.g., “mg/mg³,” “ppm

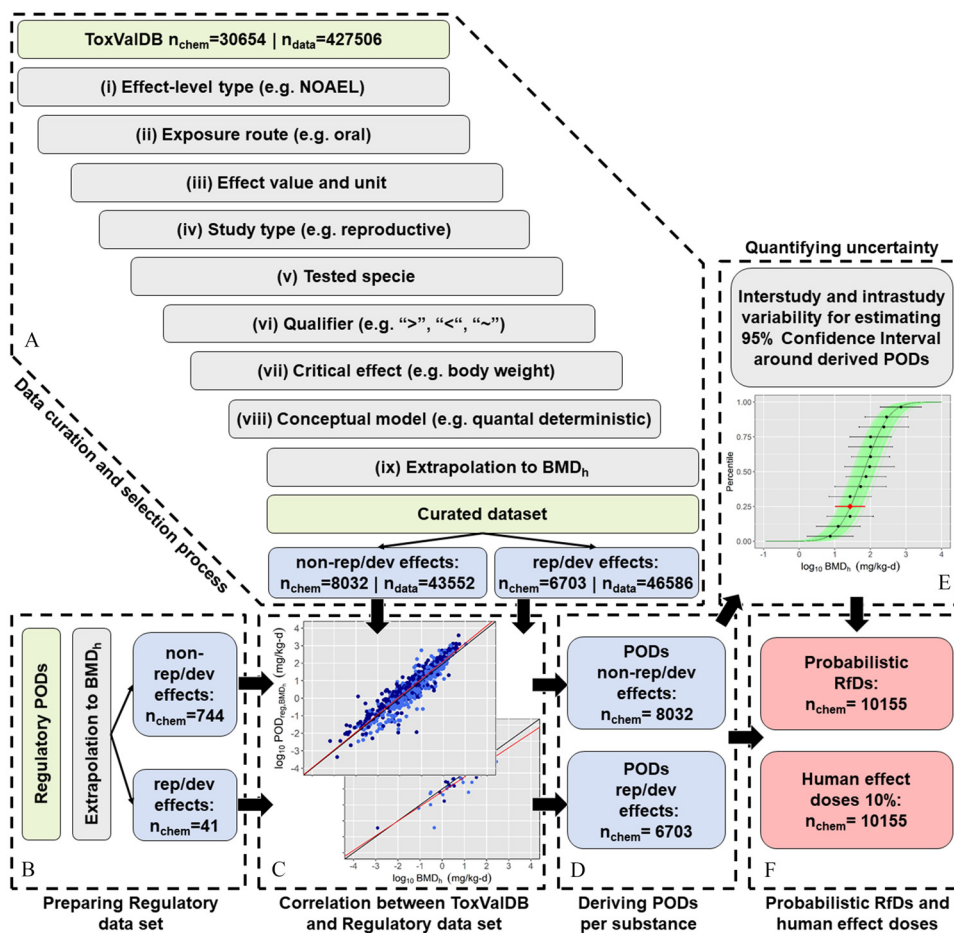


Figure 1. Overview of the workflow: (A) semiautomated data curation and selection process applied to the collected *in vivo* data from ToxValDB; (B) collection and extrapolation of regulatory PODs; (C) analysis of the correlation between ToxValDB and regulatory POD data; (D) systematic derivation of oral PODs from the curated data sets, differentiating between general noncancer (nonreproductive/developmental) and reproductive/developmental effects; (E) quantification of the substance-specific uncertainty of the derived PODs from intra- and interstudy variability; and (F) derivation of probabilistic RfD and human population effect doses ($I=10\%$). Note: BMD_h, human equivalent benchmark dose; n_{chem}, number of chemicals; n_{data}, number of data points (records); NOAEL, no observed adverse effect level; non-rep/dev, nonreproductive or developmental; POD, point of departure; rep/dev, reproductive or developmental; RfD, probabilistic reference dose; ToxValDB, Toxicity Value Database.

urine," or "mg/kg ash femur"). Specifically for records with REACH as source and "mg/kg diet" as the reported unit of effect value, we converted the reported effect values to milligrams per kilogram per day by dividing the reported effect value by 16 if the tested species was rat and by 4.5 if the tested species was mouse. Single-dose data (acute tests, unit typically in "mg/kg") were also excluded.

- Study type:** we focused on five study types: chronic, subchronic, subacute, reproductive, and developmental. Harmonization of reported study types included, for example, "fertility" being assigned to reproductive. In addition, for records with subacute or subchronic as the study type, we extrapolated their effect values to chronic by applying a subchronic-to-chronic factor of 2 and a subacute-to-chronic factor of 5.^{12,28} For records with reproductive or developmental as the study type, we did not apply any extrapolation because we assumed that the study covered the relevant window of susceptibility. The records with reported study type being different (and unconvertible) to one of the five considered were disregarded.
- Tested species:** we focused only on records providing toxicity information on mammals, excluding other species. We harmonized reported species names and grouped them into commonly tested species. For example, we grouped

records with tested species reported as mice or hamster into mouse. If no tested species were reported, we flagged the record and assumed the tested species to be rat (the predominant tested species across the retrieved data). In addition, we extrapolated the effect values of all records to humans. The interspecies body weight scaling was performed by dividing reported effect values by conversion factors (CF s) to humans estimated as follows:

$$CF = \frac{BW_h^{0.25}}{BW_a^{0.25}}$$

where BW_h is the average body weight of humans of 70 kg, and BW_a is the body weight of the tested species. As an example, by assuming an average weight for a mouse $BW_a=0.025$ kg, a $CF=7.3$ is estimated, and in case of an effect value of 10 mg/kg per day, the dose tested with a mouse is converted to an effect value for humans of 1.4 mg/kg per day.

- Qualifiers:** in cases of reported effect values accompanied by numeric qualifiers (e.g., "<," ">," "≥"), after analyzing the original sources for these records and based on expert judgment, we decided to disregard the presence of the numeric qualifiers except for NOAELs accompanied by

“<.” The effect-level types of these “NOAEL <*x*” records were converted to LOAEL given that actual effects were observed at the tested dose in the original studies.

- Critical effects:** the reported effects studied were standardized to one of the following categories: body weight, clinical chemistry, clinical signs, development, enzyme activity, food or water consumption, gross pathology, hematology, mortality/survival, multiple, neurobehavior, none, nonneoplastic histopathology, organ weight, other, reproduction, or urinalysis. In addition, for all records for which we allocated development or reproduction as critical effect category, we cross-checked this information with the previously harmonized study type category and overwrote the study type category in case of mismatch.
- Conceptual model:** based on the previously assigned standardized effect categories and study types, we assigned to each data record one of the following conceptual models: continuous, quantal-deterministic, quantal-stochastic, or multiple, following the WHO/IPCS recommendations in dose–response modeling (Table S2).^{10–12} For example, chronic records with body weight as a standardized effect category were assigned the conceptual model continuous.
- Extrapolation to benchmark dose:** based on the curated effect-level type, study type, and assigned conceptual model, we extrapolated the effect value of each record to a chronic human equivalent benchmark dose (BMD_h) based on the WHO/IPCS framework (Table S3).^{10,12} In the case of multiple possible conceptual models assigned to the same record, we calculated the BMD_h value based on the averaged results of the two assigned conceptual models. At each extrapolation step to convert the *in vivo* data to BMD_h (e.g., interspecies body weight scaling), uncertainty distributions were assigned to BMD_h. Assuming lognormal distribution for each factor, the uncertainties were combined probabilistically by applying the approximation described in the latest work in dose–response modeling.^{10–12} The probabilistically combined uncertainties quantified a total uncertainty around each extrapolated effect value (i.e., BMD_h). In Table S3, uncertainty factors are provided as the ratio between the 95th percentile and the median of the lognormal distribution (P95/P50).

After curating and applying the described semiautomated process to select the retained toxicity data from ToxValDB, we split the curated data into two distinct data sets covering general noncancer effects and reproductive/developmental effects, respectively. This repartition was performed based on each record's derived study type and critical effects (Figure 1A).

Regulatory Data

We gathered regulatory data from a previously published database of publicly available, peer-reviewed human health toxicity values reported in specific public sources, including—among others—U.S. EPA (e.g., IRIS, OPP) and California EPA (Office of Environmental Health Hazard Assessment).^{8,29} We then cross-checked these values with the November 2019 release of the U.S. EPA Regional Screening Levels (RSLs), adding additional chemicals not previously identified for which PODs could be identified.³⁰ In our study, a POD_{reg} is defined as the NOAEL, LOAEL, or BMDL associated with a reported reference dose. To ensure a consistent comparison, we extrapolated the gathered POD_{reg} to chronic human equivalent benchmark dose (POD_{reg,BMD_h}), applying the same procedure as described previously for the curated and selected ToxValDB records while also differentiating between general noncancer and reproductive/developmental effects (Figure 1B).

Comparison and Approach for Deriving Oral PODs

To systematically determine oral PODs for substances for which regulatory values were not available, we started by comparing the curated toxicity data from ToxValDB against the available POD_{reg,BMD_h} (Figure 1C). The comparison was carried out separately for general noncancer and reproductive/developmental effects for chemicals for which both POD_{reg,BMD_h} and *in vivo* data were available. For each of these substances, we assumed a lognormal distribution across BMD_h and derived a POD from the *x*-percentile of the fitted lognormal distribution (POD_{*p*xBMD_h}). The resulting POD_{*p*xBMD_h} values were then compared against the respective POD_{reg,BMD_h}. Although the curated data from ToxValDB did not necessarily cover the exact data sets used by health risk assessors to select POD_{reg}, the comparison between the resulting values informed us about the importance of potential differences.

We hypothesized that POD_{*p*xBMD_h} on the lower end of the effect values distribution was a suitable proxy for POD_{reg,BMD_h} across different chemicals.²⁰ To evaluate this hypothesis and to identify the most suitable *x*-percentile, we analyzed the correlation of POD_{reg,BMD_h} values against four different POD_{*p*xBMD_h} values, from the 5th to the 35th percentile (i.e., POD_{*p*05BMD_h}, POD_{*p*15BMD_h}, POD_{*p*25BMD_h}, and POD_{*p*35BMD_h}). In addition, to put our approach into perspective, we investigated via the Shapiro–Wilk normality test³¹ whether the BMD_h distribution for each chemical could be adequately fit by a lognormal distribution.

The two function moments used for fitting the lognormal distribution were mu (μ) and sigma (σ), which respectively denoted the log-scale population median and standard deviation of the available effect values for a substance.³² For all substances, μ was calculated from the available BMD_h. In contrast, σ was calculated from the available BMD_h only for data-rich chemicals (≥10 records available), whereas for data-poor chemicals (<10 records available), we applied a fixed standard deviation (σ_{fixed}) derived from the average across σ of data-rich chemicals. We derived two distinct σ_{fixed}, one to be applied for general noncancer effects (σ_{fixed}^{non-rep/dev}) and one for reproductive/developmental effects (σ_{fixed}^{rep/dev}). Given that the estimates of σ from <10 available records were highly unstable, we used an average shaped distribution instead of relying on the few available effect values. The derived *x*-percentile from the fitted lognormal distribution (POD_{*p*xBMD_h}) were expected to be more representative for the considered data-poor chemical.

In addition, to investigate the potential influence of remaining double entries (i.e., duplicate records) in the curated ToxValDB, we studied how much the surrogate PODs were affected in the case of keeping only records with unique derived BMD_h values, effect-level types, and tested species.

Deriving PODs per Substance

After identifying the most suitable *x*-percentile to be used as a surrogate of POD_{reg,BMD_h}, we systematically derived POD_{*p*xBMD_h} for each substance from the available records in the two curated *in vivo* data sets (Figure 1D). For a substance for which records were available in both data sets, two distinct POD_{*p*xBMD_h} values were derived separately, one for general noncancer effects (POD_{*p*xBMD_h}^{non-rep/dev}) and one for reproductive/developmental effects (POD_{*p*xBMD_h}^{rep/dev}).

Quantifying Uncertainty around the Derived PODs

To characterize the uncertainty around the derived POD_{*p*xBMD_h}, we took into account both interstudy variability and intrastudy variability (Figure 1E). These two aspects were quantified separately and expressed as the squared geometric standard deviation (GSD_{inter}²

and GSD_{intra}^2) and then combined (GSD_{total}^2)³³ to provide a 95% confidence interval (CI) for each $POD_{p \times BMD_h}$ in the two data sets:

$$GSD_{total}^2 = 10^{\sqrt{(\log_{10} GSD_{inter}^2)^2 + (\log_{10} GSD_{intra}^2)^2}}$$

GSD_{total}^2 being a unitless factor equal to $P97.5/P50$ or to $(P95/P50)^{2/1.65}$ and denoting that the distribution of 95% of all values fall within $POD_{p \times BMD_h}$ divided by GSD_{total}^2 and $POD_{p \times BMD_h}$ multiplied by GSD_{total}^2 .

We calculated GSD_{inter}^2 , which reflects the variability across available effect values, for each $POD_{p \times BMD_h}$ in the two distinct data sets. To estimate GSD_{inter}^2 , we started from the lognormal distribution fitted through the available effect values (extrapolated to BMD_h) for deriving $POD_{p \times BMD_h}$. When fitting the lognormal distribution, one of the two moments used was σ (standard deviation of the available BMD_h for a substance). We thus estimated the 95% CI of σ via the function `fitdistr` in the R package MASS,³⁴ and from this 95% CI we derived an upper and lower bound for $POD_{p \times BMD}$ ($POD_{p \times BMD_h}^{inter, upper}$ and $POD_{p \times BMD_h}^{inter, lower}$)³⁵ by fitting two new lognormal distributions using instead of σ its 95% CI. GSD_{inter}^2 was then calculated as:

$$GSD_{inter}^2 = \sqrt{POD_{p \times BMD_h}^{inter, upper} / POD_{p \times BMD_h}^{inter, lower}}$$

In contrast, GSD_{intra}^2 reflects the variability specific to the effect values. To estimate GSD_{intra}^2 , we started from the record-specific distribution around the extrapolated effect value. This record-specific distribution was based on the uncertainty distributions assigned when converting the *in vivo* data to human BMDs at the following extrapolation steps: LOAEL to NOAEL, NOAEL or BMDL to BMD, subchronic/subacute to chronic, interspecies body weight scaling and, interspecies toxicokinetics (TKs) and toxicodynamics (TDs) (Table S3).^{10–12} The record-specific uncertainty was propagated from the available records (BMD_h) to the derived $POD_{p \times BMD_h}$ via a bootstrap method. First, for each substance, 1,000 bootstrap samples were sampled from the estimated distributions around BMD_h of the available records. Second, 1,000 lognormal distributions were fitted to the bootstrap samples using μ as the median of the resampled effect values, and σ as the same σ used to derive BMD_h , based on the originally available effect values (in practice only μ varied and the same shaped distribution was always fitted to the resamples). Third, from the 1,000 fits, we derived an upper and lower bound for $POD_{p \times BMD_h}$ ($POD_{p \times BMD_h}^{intra, upper}$ and $POD_{p \times BMD_h}^{intra, lower}$). GSD_{intra}^2 was then calculated as follows:

$$GSD_{intra}^2 = \sqrt{POD_{p \times BMD_h}^{intra, upper} / POD_{p \times BMD_h}^{intra, lower}}$$

When using the derived $POD_{p \times BMD_h}$ as a surrogate of regulatory value, it was necessary to consider the additional uncertainty associated with the prediction of this regulatory value, which was obtained from the residual standard error between POD_{reg, BMD_h} and $POD_{p \times BMD_h}$. Because this residual standard error already accounted for the uncertainty on the $POD_{p \times BMD_h}$ for regulated chemicals, we took as GSD_{final}^2 the maximum between the uncertainty related to the residual standard error ($GSD_{p \times \rightarrow reg}^2$) and the substance-specific GSD_{total}^2 on the derived POD .

Deriving Probabilistic RfDs and Human Effect Doses

Following the automated workflow developed by Chiu et al.,¹² probabilistic RfDs were derived for risk assessment purposes as the lower 95% confidence bound of $HD_M^{1\%}$, that is, the daily human dose at which 1% of the population shows a level of effect M

corresponding to the effect-level type (e.g., LOAEL, NOAEL, or BMDL) reported in the database as well as the type of end point (e.g., continuous, quantal-deterministic, or quantal-stochastic) (Figure 1F). $HD_M^{1\%}$ values were calculated from the provided $POD_{p \times BMD_h}$ by dividing it by an extrapolation factor of 9.7 (P50) to account for variability in sensitivity between the median human and the first percentile human.¹⁰ The 90% CI of $HD_M^{1\%}$ was calculated combining probabilistically GSD_{final}^2 and the uncertainty factor (i.e., $P95/P50 = 4.3$) assigned to the human variability at the first percentile to yield the 90% CI of $GSD_{final}^{1.65} = (GSD_{final}^2)^{1.65}$.¹⁰ We directly implemented the approximate approach by Chiu et al.¹² given that in their study it yielded results within 20–30% of the Monte Carlo simulation. We then compared the derived lower 95% confidence bound of $HC_M^{1\%}$ against the related regulatory RfD (if available) to investigate the potential influence of the database uncertainty factor (UF_d). This factor accounts for data gaps and is typically equal to 1, 3, and 10 as a function of the data coverage for different end points.³⁶ UF_d was applied when deriving regulatory RfDs but it was not directly included in the WHO/IPCS framework.¹² This helped us understand whether the derived toxicity values were consistent with regulatory RfDs and identify potential biases. To put the obtained results into perspective, the derived probabilistic RfDs were finally compared against the population median chemical intake rates provided by the Systematic Empirical Evaluation of Models (SEEM) meta-model.³⁷

For LCIA purposes, we derived effect doses at which 10% of the population shows a level of effect M ($HD_M^{10\%}$). $HD_M^{10\%}$ was derived from the provided $POD_{p \times BMD_h}$ by dividing it by 3.49 (P50) as an extrapolation factor to account for the human variability between the 50% and the 10% incidence level.¹⁰ $HD_M^{10\%}$ -related uncertainty was calculated by combining probabilistically GSD_{final}^2 of $POD_{p \times BMD_h}$ and the uncertainty factor assigned to the human variability at the 10th percentile, that is, $P97.5/P50 = 2.67$,¹⁰ $HD_M^{10\%}$ being also defined as ED10 by Fantke et al.¹

Data Analysis

The curation and selection process of the toxicity data and all the analyses were carried out using the open source statistical software R (version 3.6.1; R Development Core Team). All figures were generated by ggplot2 package³⁸ in R. The R code for deriving PODs from the curated data sets is available in the Supplemental Material, “R code for deriving points of departure from the curated datasets.”

Results

Curated Toxicity Test Data Sets

After the application of the semiautomated data curation and selection process, we obtained two distinct data sets, the first covering general noncancer effects composed of $n = 43,528$ records and providing toxicity information for $n = 8,023$ substances, the second covering reproductive/developmental effects composed of $n = 46,565$ records for $n = 6,697$ substances. The fraction of records excluded at each step of the curation and selection process is provided in Table S1. Table S4 presents the summary statistics of the two curated data sets, and Figure 2A,B visualizes the effect values (all extrapolated to BMD_h) distribution across curated records and the underlying effect-level types and study types information. Most of the records had NOAEL as effect-level type in both data sets, 71% ($n = 31,082$) for the general noncancer effects data set (Figure 2A) and 78% ($n = 36,381$) for the reproductive/developmental effects data set (Figure 2B). Only a small share of the available records reported BMDL as an effect-level type ($\leq 1\%$, $n = 581$), the rest of the data being reported as LOAEL. In the general noncancer data set, 33%

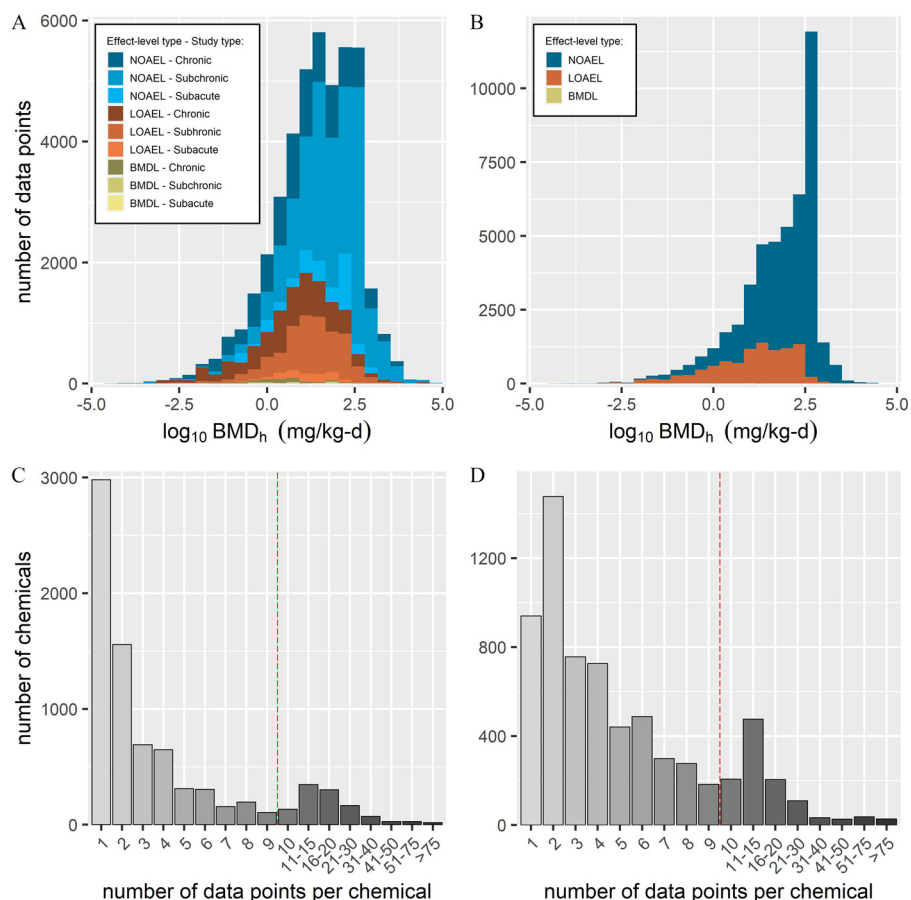


Figure 2. Distribution across curated records of (A) the effect values (BMD_h) and the underlying effect-level and study types for the general noncancer effects data set ($n = 43,528$) and for (B) the reproductive/developmental effects data set ($n = 46,565$), and number of available records for each chemical in (C) the general noncancer effects data set ($n = 43,528$) and in (D) the reproductive/developmental effects data set ($n = 46,565$). The red dashed lines in (C) and (D) divide data-poor chemicals (< 10 records available) and data-rich chemicals (≥ 10 records available). Corresponding numeric data for (A) and (C) are available in Excel Table S1, and for (B) and (D) in Excel Table S2. Note: BMD_h , chronic human equivalent benchmark dose; BMDL, benchmark dose lower confidence limit; LOAEL, lowest observed adverse effect level; NOAEL, no observed adverse effect level.

($n = 14,605$) of the records were reported as chronic, a majority of 58% ($n = 25,342$) as subchronic, and only 8% ($n = 3,581$) as subacute.

In both data sets, BMD_h ranged substantially across records by > 10 orders of magnitude. More specifically, for general noncancer effects, BMD_h ranged from 6×10^{-9} to 2.2×10^5 mg/kg per day, with a median value of 31 mg/kg per day, and for reproductive/developmental effects from 7.3×10^{-10} to 2.6×10^5 mg/kg per day, with a median value of 100 mg/kg per day. Figure 2C,D shows the number of chemicals falling within different bins of reported data points per chemical, differentiating between the two curated data sets. We observed that only a limited number of records were available for the majority of the substances. For example, 47% ($n = 3,169$) of the chemicals in the reproductive/developmental effects data set had less than four available records (Figure 2D).

The rat was the most commonly reported tested species in both data sets, followed by the mouse. Together, these two tested species represented $> 80\%$ ($n = 76,548$) of the reported tested species across the two data sets. The third most common tested species were dog in the general noncancer effects data set and rabbit in the reproductive/developmental effects data set. The tested species was not reported for 5% ($n = 3,827$) of the records across both data sets; we flagged these records and assumed rats as the tested species (Figure S2).

Figure 3 presents the effect values (BMD_h), related effect-level types, and POD_{reg,BMD_h} (when available) for all the chemicals

covered in the general noncancer effects data set (Figure 3A) and the reproductive/developmental effects data set (Figure 3B). For a given chemical, the observed variability in BMD_h spanned up to 7 orders of magnitude, and across chemicals, we observed an average standard deviation of half an order of magnitude. As a general trend, we observed that POD_{reg,BMD_h} fell on the lower half of the effect values distribution across different chemicals. In addition, based on the Shapiro–Wilk tests carried out, we found that the BMD_h distribution for each chemical could be adequately fit by a lognormal distribution, with $p > 0.05$ for the majority of the chemicals in the two data sets.

The $n = 90,093$ curated and selected toxicity records from the ToxValDB are provided in the Supplemental Material, differentiating between the general noncancer effects data set (Excel Table S1) and the reproductive/developmental effects data set (Excel Table S2). POD_{reg,BMD_h} extrapolated to chronic human equivalent benchmark doses are available for $n = 744$ chemicals for general noncancer effects (Excel Table S3) and for $n = 41$ chemicals for reproductive/developmental effects (Excel Table S4).

Comparison with Regulatory Toxicity Values

To characterize the distribution of the toxicity values for data-rich chemicals with at least 10 records available, we directly used the available effect values (BMD_h) to derive a chemical-specific standard deviation given that the available records were sufficient to represent

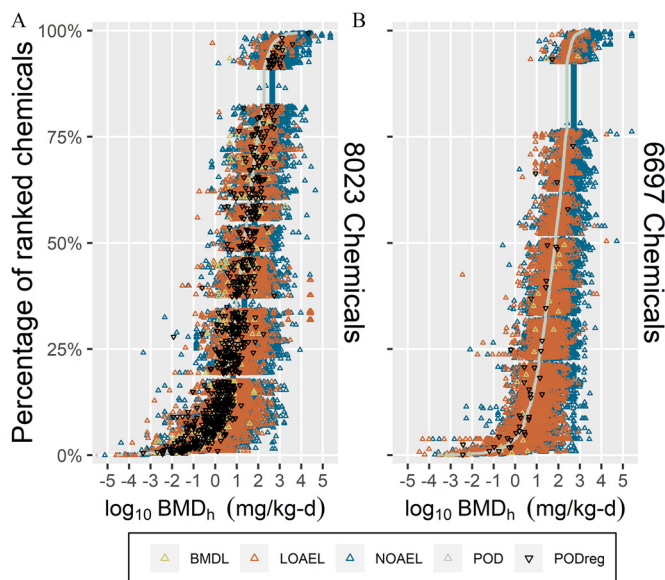


Figure 3. Curated effect values (extrapolated to BMD_h), their underlying effect-level types, the corresponding regulatory PODs (POD_{reg,BMD_h}) and derived POD_{p25BMD_h} (gray data points) for each of the chemicals covered by the *in vivo* data, differentiating between (A) general noncancer effects and (B) reproductive/developmental effects. Chemicals are ranked by derived PODs, the gray curve representing the percentage of chemicals above a certain POD. Corresponding numeric data for NOAELs, LOAELs, and BMDLs in (A) and (B) are available in Excel Table S1 and Excel Table S2, respectively; corresponding numeric data for regulatory PODs in (A) and (B) are available in Excel Table S3 and Excel Table S4, respectively; corresponding numeric data for derived PODs in (A) and (B) are available in Excel Table S5. Note: BMD_h , chronic human equivalent benchmark dose; BMDL, benchmark dose lower confidence limit; LOAEL, lowest observed adverse effect level; NOAEL, no observed adverse effect level; POD, point of departure; POD_{p25BMD_h} , point of departure derived from the 25th percentile of the fitted lognormal distribution to the curated effect values extrapolated to chronic human equivalent benchmark dose; POD_{reg,BMD_h} , point of departure associated with a reported reference dose extrapolated to chronic human equivalent benchmark dose.

and cover different potential effects. We also derived average standard deviations across data-rich chemicals of $\log_{10} \sigma_{fixed}^{non-rep/dev} = 0.55$ for general noncancer effects and $\log_{10} \sigma_{fixed}^{rep/dev} = 0.45$ for reproductive/developmental effects (Figure S3). We then applied these

averages to all data-poor chemicals with <10 records, for which chemical-specific σ would not be reliable.

Using these standard deviations, we constructed lognormal distributions of BMD_h and compared four different $POD_{p\%BMD_h}$ (i.e., 5th, 15th, 25th, and 35th percentiles) to the curated POD_{reg,BMD_h} values. This analysis identified the 25th percentile (i.e., POD_{p25BMD_h}) as the best approximation of POD_{reg,BMD_h} for both effect data sets (Figure S4). Figure 4 compares the estimated POD_{p25BMD_h} and the respective POD_{reg,BMD_h} for general noncancer effects (Figure 4A) and for reproductive/developmental effects (Figure 4B). For both data sets, the estimated POD_{p25BMD_h} correlated well with the available POD_{reg,BMD_h} , with a coefficient of determination $R^2 \geq 0.78$ and a residual standard error (RSE) ≤ 0.53 of the log-transformed values. In addition, we investigated the few outliers present in Figure 4 and did not identify specific trends or clusters. These outliers covered a large chemical space and included metals, insecticides, and phthalate plasticizers. Hence, no chemical categories appeared to be more problematic than others.

Recommended PODs

After identifying the 25th percentile of the distribution as the best approximation of POD_{reg,BMD_h} , we derived surrogate PODs (i.e., POD_{p25BMD_h}) for $n = 10,145$ substances. More specifically, from the general noncancer effects data set, we derived surrogate PODs for $n = 8,023$ substances, and from the reproductive/developmental effects data set, we derived surrogate PODs for $n = 6,697$ substances. For each of the $n = 4,575$ substances common to both data sets, two distinct surrogate PODs were thus derived, one from each data set.

Figure 3 presents the derived surrogate PODs ranked in increasing order (dark gray curve), together with the underlying BMD_h , as well as the related POD_{reg,BMD_h} where available. The derived surrogate POD values range across chemical substances >8 –10 orders of magnitude, from 3.1×10^{-6} to 1.1×10^4 mg/kg per day, with a median value of 22 mg/kg per day for general noncancer effects, and from 2.8×10^{-4} to 1.3×10^4 mg/kg per day, with a median value of 76 mg/kg per day for reproductive/developmental effects. Examples of substances with the lowest POD estimates (i.e., highest potential toxicity) in both data sets include dioxins [e.g., 2,3,7,8-tetrachlorodibenzo-p-dioxin, Chemical Abstract Service (CAS): 1746-01-6], polychlorinated dibenzofurans (e.g., 2,3,4,7,8-pentachloro-dibenzofuran, CAS: 57117-31-4) and, heavy metals (e.g.,

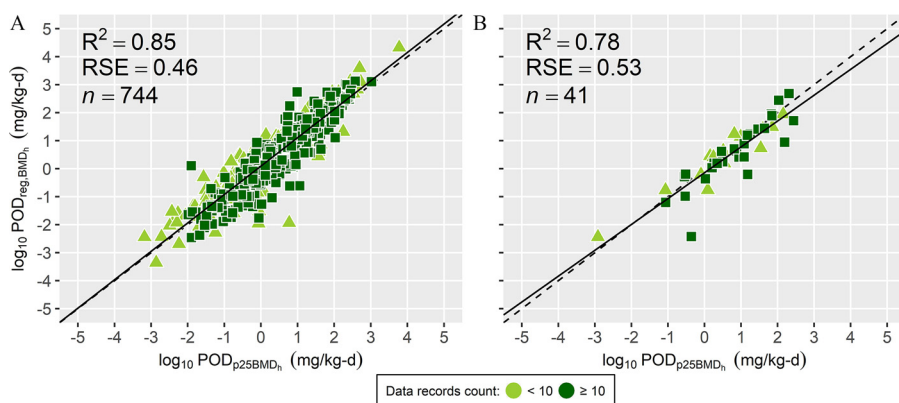


Figure 4. Comparison between estimated POD_{p25BMD_h} and available regulatory POD values (POD_{reg,BMD_h}) for data-rich (dark green rectangle, ≥ 10 records available) and data-poor chemicals (light green triangle, <10 records available), differentiating between (A) general noncancer effects and (B) reproductive/developmental effects. The dashed line represents the 1:1 line, and the solid line represents the best fit. Corresponding numeric data for POD_{reg,BMD_h} in (A) and (B) are available in Excel Table S3 and Excel Table S4, respectively; corresponding numeric data for POD_{p25BMD_h} is available in Excel Table S5. Note: POD, point of departure; POD_{p25BMD_h} , point of departure derived from the 25th percentile of the fitted lognormal distribution to the curated effect values extrapolated to chronic human equivalent benchmark dose; POD_{reg,BMD_h} , point of departure associated with a reported reference dose extrapolated to chronic human equivalent benchmark dose; RSE, residual standard error.

lead, CAS: 7439-92-1). The gaps observed in Figure 3 were linked to those substances for which only one record was available with original reported effect values corresponding to standard tested dosimetry doses (e.g., 10, 100, 1,000 mg/kg per day). Excel Table S5 provides all derived PODs and the number of underlying effect values.

We compared the derived POD values for the $n=4,575$ substances with two distinct surrogate PODs derived from each data set. As a general trend, we observed that the higher the toxicity for general noncancer, the higher for reproductive/developmental effects. However, we also observed outliers. For the same chemical, PODs covering general noncancer effects were lower (i.e., higher toxicity) than the ones covering reproductive/developmental effects by a median factor of 2. We also observed a high variability of the ratio of the two POD values across chemicals, going from a factor of 0.0003 to 4,000 (Figure S5). In addition, we investigated the potential influence of duplicate records in the curated data set when deriving POD values. With this analysis, we found that the difference of POD values was limited for the majority of the substances, with an average increase of only 7%, with no greater than a factor 2 increase for 95% ($n=4,806$) of substances for both general noncancer and reproductive/developmental effects (Excel Table S6).

Concerning regulatory values, POD_{reg,BMD_h} were available for both general noncancer effects and reproductive/developmental effects for 23 chemicals, of which 15 were pesticides with extensive testing requirements. The range of derived POD values for these 23 chemicals spanned almost 5 orders of magnitude (from 2.4 to 1,024 mg/kg per day), but the ratios between the two POD values across chemicals were all less than 1 order of magnitude, which corresponded to the uncertainty in the POD from any one study. The results of this comparison are given in Excel Table S7.

Uncertainty Estimates for PODs

To derive a 95% CI around the derived PODs (POD_{p25BMD_h}), we first estimated the two types of uncertainty for each POD, namely GSD_{inter}^2 and GSD_{intra}^2 , reflecting interstudy and intrastudy variability (Figure S6).

In the two data sets, GSD_{inter}^2 of data-rich chemicals increased with the corresponding σ of the available BMD_h while decreasing with the number of data points, from a maximum value of a factor $GSD_{inter}^2=3.1$, down to a factor <1.2 with >100 data points (Figure S7A,B). For data-poor chemicals (<10 records available), we assigned a fixed $GSD_{inter}^2=2.4$, calculated as the 97.5th percentile of the estimated GSD_{inter}^2 across substances with $n=10$ records available, given that GSD_{inter}^2 might be unreliable and highly biased by the limited number of effect values available.

For intrastudy variability, GSD_{intra}^2 values were estimated via the 1,000-bootstrap-samples approach across PODs in the two data sets and ranged from a factor of 1.1 to a factor of 14.4 (Figure S7C,D). For chemicals with a single record available, we defined $GSD_{intra_single}^2$ as the upper-bound of the estimated GSD_{intra}^2 across substances with two records available, differentiating between general noncancer effects ($GSD_{intra_single}^2=15$) and reproductive/developmental effects ($GSD_{intra_single}^2=12$) (Figure S7C,D).

Finally, we combined these two uncertainties to characterize an overall substance-specific GSD_{total}^2 for each derived POD. Even though there was variability in GSD_{total}^2 across substances with the same number of records due to differences in the variability of the underlying data, this variability systematically decreased with the increase in the number of records available (Figure S7). When comparing with regulatory values, the uncertainty factor of $GSD_{p25\rightarrow reg}^2=10^{1.96\times 0.46}=8$ for general noncancer and $GSD_{p25\rightarrow reg}^2=10^{2.02\times 0.53}=12$ for reproductive/

developmental effects were also considered to reflect the use of POD_{p25BMD} as a suitable approximation of POD_{reg} . We took as the final substance-specific GSD_{final}^2 the maximum between GSD_{total}^2 and $GSD_{p25\rightarrow reg}^2$. Estimated GSD_{final}^2 ranged from $GSD_{final}^2=8$ up to $GSD_{total}^2=17.2$ for general noncancer effects, and up to $GSD_{final}^2=13.9$ for reproductive/developmental effects. The distributions of the resulting surrogate PODs (POD_{p25BMD_h}) with their characterized 95% CIs are displayed in Figure S8.

In addition, we investigated for which fraction of substances the available regulatory PODs (POD_{reg,BMD_h}) were falling within the 95% CI of the derived surrogate PODs to put the provided results in perspective. From this analysis, we observed that for the majority of the considered chemicals, POD_{reg,BMD_h} were well within the estimated 95% CI, which corresponded to 707 of 744 chemicals for general noncancer effects and 3 of 41 for reproductive/developmental effects (Figure S9).

Probabilistic RfDs and Human Effect Doses

Starting from the recommended PODs, we first derived probabilistic RfDs as the lower 95% confidence bound of $HD_M^{1\%}$, using the WHO/IPCS framework. Because this framework focuses on end point-specific uncertainties and RfDs, an additional database uncertainty factor (UF_d) needed to be included when deriving probabilistic RfDs comparable to and consistent with regulatory RfDs.

To derive probabilistic RfDs, the following additional UF_d were thus applied: The lower 95% confidence bound of $HC_M^{1\%}$ was divided by $UF_d=10$ for substances with very poor data availability ($n\leq 3$ records), by $UF_d=3$ for substances with intermediary data availability ($3<n<10$ records), and by $UF_d=1$ for data-rich substances ($n\geq 10$ records). For data-rich chemicals, the probabilistic RfD value was thus equal to the lower 95% confidence bound of $HD_M^{1\%}$. The derived probabilistic RfDs showed a good correlation with the regulatory RfDs, with a $R^2=0.58$ and $RSE=0.79$ evaluated on log-scale for the 1:1 line (Figure S10B). In contrast, neglecting UF_d would lead to a systematic overestimation of the RfDs (Figure S10A; $R^2=0.54$, $RSE=0.82$).

Derived probabilistic RfDs were on average lower than surrogate PODs by a factor of 800 and ranged across chemicals by 8–10 orders of magnitude, with a median value of 0.04 mg/kg per day for general noncancer effects (Figure 5A) and 0.1 mg/kg per day for reproductive/developmental effects (Figure 5B). The derived probabilistic RfDs could then be used to put exposures into perspective by comparing them with the population median chemical intake rates and their upper 95% confidence bound estimated via the SEEM meta-model. This analysis highlighted that only for $n=14$ chemicals, the best estimate of the median intake rates were higher than derived probabilistic RfDs. In contrast, when considering the upper 95% confidence bound, median intake rates were higher than derived probabilistic RfDs for $\sim 23\%$ ($n=1,127$) of the substances for which SEEM intake rates were available (Figure 5), substances that might deserve further scrutiny in priority.

Second, from the surrogate PODs we derived best estimates of human population effect doses 10% ($HD_M^{10\%}$) following the WHO/IPCS framework and the latest recommendations for deriving human dose–response factors for noncancer end points for LCIA. The derived $HD_M^{10\%}$ ranged across chemicals by 8–10 orders of magnitude, with a median value of 6.3 mg/kg per day for general noncancer effects and 21.8 mg/kg per day for reproductive/developmental effects. In both data sets, the characterized uncertainties of $HD_M^{10\%}$ (i.e., 95% CI) were on average equal to a factor of 14 and spanned up to a factor of 20.3 (Figure S11). Excel Table S5 provides the derived probabilistic RfDs and $HD_M^{10\%}$

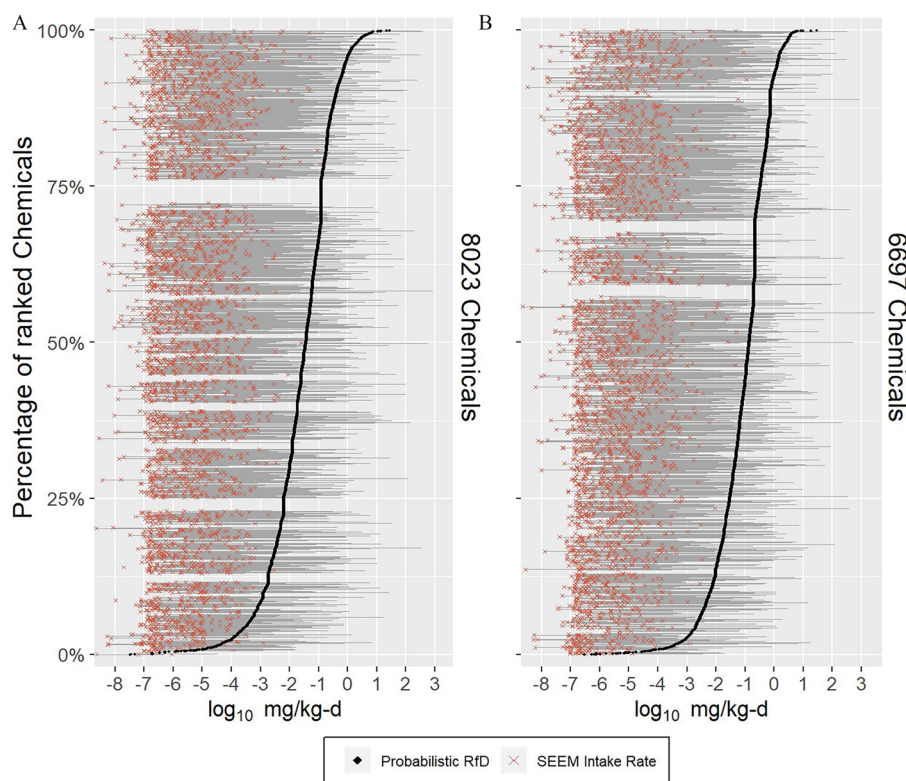


Figure 5. Derived probabilistic reference doses (RfD = lower 95% confidence bound of $HD_M^{1\%}$) and population median chemical intake rates, differentiating between (A) general noncancer effects and (B) reproductive/developmental effects. Substances are ranked in increasing order based on the derived probabilistic RfDs. The upper 95% confidence bound of the SEEM Intake rates (error bars) reflects uncertainty around the population median intake rate and does not reflect population variability. Corresponding numeric data for probabilistic RfDs are available in Excel Table S5. Note: $HD_M^{1\%}$, the daily human dose at which 1% of the population shows a level of effect M corresponding to the effect-level type reported in the database and the end point type; SEEM, Systematic Empirical Evaluation of Models.

with related uncertainties, and Excel Table S8 provides an example of their calculation from two records for an arbitrary substance.

Discussion

Applicability of the Derived Toxicity Values

By applying the presented semiautomated curation and extrapolation approach, we provided PODs consistent with regulatory values for >10,000 substances, substantially expanding the chemicals coverage for which toxicity values could be derived, from 744 to 8,023 chemicals for general noncancer effects, and in an even higher proportion, from 41 to 6,697 chemicals for reproductive/developmental effects. The derived probabilistic RfDs and $HD_M^{10\%}$ can be used in a variety of chemical management and exposure and impact assessment frameworks, including the evaluation of human toxicity impacts in LCIA,^{1,23,24} ranking and prioritization of chemicals for additional study and evaluation, chemical safety and risk management, as well as alternatives assessment for chemical substitution. By increasing the coverage of chemical substances for human toxicity effect modeling, our results also fill a critical gap in toxicity information availability highlighted, for example, in recent high-throughput exposure and risk screening studies.^{39–41} Indeed, even though exposure estimates were quantified for hundreds of different substances in such studies, the lack of toxicity data prevented a comprehensive risk evaluation for all the studied chemicals.

In addition, the derived PODs and related $HD_M^{10\%}$ are following the globally recommended approach for deriving health effect factors for noncancer end points by differentiating between general noncancer effects and reproductive/developmental effects, enabling us to then account for the average ~20-fold highest severity

of reproductive/development effect when evaluating disability adjusted life years.^{1,22} At the same time, the proposed approach is analogous to current practices of environmental risk assessment and LCIA for ecotoxicity characterization,⁴² where species sensitivity distributions are derived by fitting a lognormal distribution to effect values to quantify critical effect levels and related impacts in ecosystems.³²

Finally, by estimating oral doses as surrogates of regulatory values also for data-poor chemicals (<10 records), our proposed approach potentially helps in reducing costs and time (as well as ethical concerns) related to using large numbers of animals to derive a complete set of toxicity studies covering different effects. Although our approach cannot substitute for the rigorous health assessments of chemicals potentially of concern, nevertheless, it might support the work of health risk assessors at multiple levels for screening purposes when a chemical of concern has not yet been thoroughly tested or reviewed.⁸ Most importantly, our approach provides a reliable alternative wherever regulatory toxicity values are absent but where other subacute, subchronic, or chronic toxicity data are available.

Limitations of the Proposed Approach

Our proposed approach also comes with limitations. First, we also derived PODs for many data-poor chemicals (<10 records) and so potentially missed critical effects not covered by the considered studies, thus underestimating the actual chemical toxicity. We addressed this limitation by assigning a fixed value of the standard deviation derived from the set of chemicals with sufficient reported data records to substances with <10 available effect values, thus fitting a lognormal distribution with a predefined average shape. This higher uncertainty for data-poor chemicals is reflected via the

high GSD_{total}^2 values and related 95% CIs on the reported PODs, which are highly dependent on the number of effect values available for fitting the distribution, that is, the lower the number of records, the higher the GSD_{total}^2 (Figure S7). For a given chemical, the observed average standard deviation of half an order of magnitude for BMD_h values could arise from different sources, including different critical effects studied, different species tested in different environmental conditions (i.e., biological variability), as well as systematic errors (e.g., measurement errors, different experimental protocols, measurement tools).^{3,43,44} However, given that the rat was the most commonly tested species, the observed variability was rather associated with the different effects studied in addition to the intrinsic variability in measuring toxicity effects.

Second, specifically for reproductive/developmental effects, the comparison against POD_{reg,BMD_h} was carried out for only $n = 41$ substances. Thus, the choice of selecting the 25th percentile (POD_{p25BMD_h}) of the fitted lognormal distribution to the available effect values is less reliable than for general noncancer effects, for which regulatory data were available for $n = 744$ chemicals. We made this choice for consistency and, considering that, we still observed a good correlation compared with the other percentiles tested. Nevertheless, increasing the pool of regulatory data covering reproductive/developmental effects is urgently needed to verify that POD_{p25BMD_h} would still be the best approximation of POD_{reg,BMD_h} also for these effects.

Third, there are possible remaining double entries (i.e., duplicate records) in the retrieved ToxValDB. More specifically, these potential double entries are due to the fact that ToxValDB is collecting experimental toxicity data from >40 publicly available sources, which in turn are also gathering experimental toxicity data from different sources, as well as running actual experimental tests. Therefore, there is the risk that in ToxValDB for the same substance, various records from different sources might be available but reporting the same results from a given experimental test. Based on testing the potential influence of keeping only records with unique derived BMD_h values, effect-level types, and tested species on our results, we found that the difference in POD values was substantial only for a small fraction of substances for both general noncancer and reproductive/developmental effects. Furthermore, the presence of duplicates is already accounted for statistically through the choice of the 25th percentile to represent the surrogate POD given that this percentile was derived using data sets that may have included duplicates.

Finally, there is a limit on how accurately a toxicity value can be predicted, and this is an intrinsic limitation of our approach and of any other approach that uses reported toxicity test data as a starting point. This is because risk estimates can vary widely across regulatory settings, even for the same chemical, despite the same underlying toxicity data set and the rigorous scientific judgment involved in developing toxicity values.^{8,45} Nevertheless, our results suggest that using the 25th percentile (POD_{p25BMD_h}) of the fitted lognormal distribution to the available effect values for a substance is an efficient method for estimating a POD that would be selected in a regulatory context.

Future Research Needs

To further advance this effort toward using experimental animal data to derive PODs for human toxicity effects, future research needs include extending the current approach to cover additional exposure routes, such as inhalation and dermal exposure, given that we focused our work on oral toxicity owing to the higher data availability than for other exposure routes. Covering additional exposure routes is crucial, especially for exposure and impact assessment frameworks aiming at comparing chemicals impacts across exposure routes.^{39,46} Similarly, in our study, we

differentiate between PODs for reproductive/developmental effects and general noncancer effects owing to the difference in severity of these two disease categories to affect human lifetime loss.^{1,22} Nevertheless, future work should focus on increasing this differentiation and providing more critical effect-specific PODs, such as endocrine disruption effects.⁴⁷

Conclusions

Given the large number of new and existing substances requiring assessment, there is a pressing need for cost-effective and rapid nonanimal alternatives,⁴⁸ which is in line with the need for a transition toward more sustainable chemistries^{49,50} through the use of novel and innovative digitalization methods.⁵¹ Such methods will facilitate a broader coverage of chemicals that can be considered in a rapid screening, quantitative assessments of chemical emissions, along with product life cycles, chemical substitution, and risk prioritization. Our proposed surrogate PODs, probabilistic RfDs, and $HD_M^{10\%}$ constitute a valuable starting point for addressing these needs for substances lacking regulatory assessments.

Acknowledgments

We thank Y. Emará (Technical University of Denmark) for the discussion of the method for quantifying the uncertainty around points of departure. This research was funded in part, by grants P42 ES027704 and P30 ES029067 from the National Institute of Environmental Health Sciences. This work was supported by the Global Best Practices on Emerging Chemical Policy Issues of Concern under the UN Environment's Strategic Approach to International Chemicals Management (SAICM; GEF project 9771, grant S1-32GFL-000632), by the Safe and Efficient Chemistry by Design (SafeChem) project funded by the Swedish Foundation for Strategic Environmental Research (grant DIA 2018/11), and by the Partnership for the Assessment of Risks from Chemicals (PARC) project (grant 101057014) funded under the European Union's Horizon Europe Research and Innovation program.

References

1. Fantke P, Chiu WA, Aylward L, Judson R, Huang L, Jang S, et al. 2021. Exposure and toxicity characterization of chemical emissions and chemicals in products: global recommendations and implementation in USEtox. *Int J Life Cycle Assess* 26(5):899–915, PMID: 34140756, <https://doi.org/10.1007/s11367-021-01889-y>.
2. Fantke P, Huang L, Overcash M, Griffing E, Jolliet O. 2020. Life cycle based alternatives assessment (LCAA) for chemical substitution. *Green Chem* 22(18):6008–6024, <https://doi.org/10.1039/D0GC01544J>.
3. Pradeep P, Friedman KP, Judson R. 2020. Structure-based QSAR models to predict repeat dose toxicity points of departure. *Comput Toxicol* 16:100139, PMID: 34017928, <https://doi.org/10.1016/j.comtox.2020.100139>.
4. Pham LL, Watford SM, Pradeep P, Martin MT, Thomas RS, Judson RS, et al. 2020. Variability in *in vivo* studies: defining the upper limit of performance for predictions of systemic effect levels. *Comput Toxicol* 15:100126, PMID: 33426408, <https://doi.org/10.1016/j.comtox.2020.100126>.
5. Jolliet O, Huang L, Hou P, Fantke P. 2021. High throughput risk and impact screening of chemicals in consumer products. *Risk Anal* 41(4):627–644, PMID: 33073419, <https://doi.org/10.1111/risa.13604>.
6. Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, et al. 2009. The toxicity data landscape for environmental chemicals. *Environ Health Perspect* 117(5):685–695, PMID: 19479008, <https://doi.org/10.1289/ehp.0800168>.
7. Wang Z, Walker GW, Muir DCG, Nagatani-Yoshida K. 2020. Toward a global understanding of chemical pollution: a first comprehensive analysis of national and regional chemical inventories. *Environ Sci Technol* 54(5):2575–2584, PMID: 31968937, <https://doi.org/10.1021/acs.est.9b06379>.
8. Wignall JA, Muratov E, Sedykh A, Guyton KZ, Tropsha A, Rusyn I, et al. 2018. Conditional toxicity value (CTV) predictor: an *in silico* approach for generating quantitative risk estimates for chemicals. *Environ Health Perspect* 126(5):057008, PMID: 29847084, <https://doi.org/10.1289/EHP2998>.
9. Chiu WA, Paoli GM. 2021. Recent advances in probabilistic dose–response assessment to inform risk-based decision making. *Risk Anal* 41(4):596–609, PMID: 32966629, <https://doi.org/10.1111/risa.13595>.

10. WHO/IPCS (World Health Organization and International Programme on Chemical Safety). 2018. *Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization*, 2nd ed. <https://apps.who.int/iris/handle/10665/259858> [accessed 17 March 2020].
11. Chiu WA, Slob W. 2015. A unified probabilistic framework for dose–response assessment of human health effects. *Environ Health Perspect* 123(12):1241–1254, PMID: 26006063, <https://doi.org/10.1289/ehp.1409385>.
12. Chiu WA, Axelrad DA, Dalajamts C, Dockins C, Shao K, Shapiro AJ, et al. 2018. Beyond the RfD: broad application of a probabilistic approach to improve chemical dose–response assessments for noncancer effects. *Environ Health Perspect* 126(6):067009, PMID: 29968566, <https://doi.org/10.1289/EHP3368>.
13. Judson R. 2019. *ToxValDB: Compiling Publicly Available In Vivo Toxicity Data*. The United States Environmental Protection Agency's Center for Computational Toxicology and Exposure. Presentation. <https://doi.org/10.23645/epacomptox.7800653.v1>.
14. Li L, Zhang Z, Men Y, Baskaran S, Sangion A, Wang S, et al. 2022. Retrieval, selection, and evaluation of chemical property data for assessments of chemical emissions, fate, hazard, exposure, and risks. *ACS Environ Au* 2(5):376–395, <https://doi.org/10.1021/acsenvironau.2c00010>.
15. Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ Health Perspect* 121(1):23–31, PMID: 23086705, <https://doi.org/10.1289/ehp.1205687>.
16. Aurisano N, Fantke P. 2022. Semi-automated harmonization and selection of chemical data for risk and impact assessment. *Chemosphere* 302:134886, PMID: 35537623, <https://doi.org/10.1016/j.chemosphere.2022.134886>.
17. Aurisano N, Albizzati PF, Hauschild M, Fantke P. 2019. Extrapolation factors for characterizing freshwater ecotoxicity effects. *Environ Toxicol Chem* 38(11):2568–2582, PMID: 31393623, <https://doi.org/10.1002/etc.4564>.
18. Smith MN, Cohen Hubal EA, Faustman EM. 2020. A case study on the utility of predictive toxicology tools in alternatives assessments for hazardous chemicals in children's consumer products. *J Expo Sci Environ Epidemiol* 30(1):160–170, PMID: 31501452, <https://doi.org/10.1038/s41370-019-0165-y>.
19. Aurisano N, Weber R, Fantke P. 2021. Enabling a circular economy for chemicals in plastics. *Curr Opin Green Sustain Chem* 31:100513, <https://doi.org/10.1016/j.cogsc.2021.100513>.
20. Friedman KP, Gagne M, Loo LH, Karamertzanis P, Netzeva T, Sobanski T, et al. 2020. Utility of *in vitro* bioactivity as a lower bound estimate of *in vivo* adverse effect levels and in risk-based prioritization. *Toxicol Sci* 173(1):202–225, PMID: 31532525, <https://doi.org/10.1093/toxsci/kfz201>.
21. Fantke P, Aurisano N, Provoost J, Karamertzanis PG, Hauschild M. 2020. Toward effective use of REACH data for science and policy. *Environ Int* 135:105336, PMID: 31884133, <https://doi.org/10.1016/j.envint.2019.105336>.
22. Huijbregts MAJ, Rombouts LJA, Ragas AMJ, van de Meent D. 2005. Human-toxicological effect and damage factors of carcinogenic and noncarcinogenic chemicals for life cycle impact assessment. *Integr Environ Assess Manag* 1(3):181–244, PMID: 16639884, <https://doi.org/10.1897/2004-007R.1>.
23. Jolliet O, Ernstoff AS, Csiszar SA, Fantke P. 2015. Defining product intake fraction to quantify and compare exposure to consumer products. *Environ Sci Technol* 49(15):8924–8931, PMID: 26102159, <https://doi.org/10.1021/acs.est.5b01083>.
24. Fantke P, Ernstoff AS, Huang L, Csiszar SA, Jolliet O. 2016. Coupled near-field and far-field exposure assessment framework for chemicals in consumer products. *Environ Int* 94:508–518, PMID: 27318619, <https://doi.org/10.1016/j.envint.2016.06.010>.
25. Martin MT, Judson RS, Reif DM, Kavlock RJ, Dix DJ. 2009. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. *Environ Health Perspect* 117(3):392–399, PMID: 19337514, <https://doi.org/10.1289/ehp.0800074>.
26. Watford S, Pham LL, Wignall J, Shin R, Martin MT, Friedman KP. 2019. ToxRefDB version 2.0: improved utility for predictive and retrospective toxicology analyses. *Reprod Toxicol* 89:145–158, PMID: 31340180, <https://doi.org/10.1016/j.reprotox.2019.07.012>.
27. Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, et al. 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J Cheminform* 9(1):61, PMID: 29185060, <https://doi.org/10.1186/s13321-017-0247-6>.
28. Guth S, Roth A, Engeli B, Lachenmeier DW, Cartus AT, Hüser S, et al. 2020. Comparison of points of departure between subchronic and chronic toxicity studies on food additives, food contaminants and natural food constituents. *Food Chem Toxicol* 146:111784, PMID: 32998026, <https://doi.org/10.1016/j.fct.2020.111784>.
29. Wignall JA, Shapiro AJ, Wright FA, Woodruff TJ, Chiu WA, Guyton KZ, et al. 2014. Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. *Environ Health Perspect* 122(5):499–505, PMID: 24569956, <https://doi.org/10.1289/ehp.1307539>.
30. U.S. EPA (U.S. Environmental Protection Agency). 2019. Regional Screening Levels (RSLs) - Generic Tables. <https://www.epa.gov/risk/regional-screening-levels-rsls-generic-tables> [accessed 17 March 2020].
31. Shapiro SS, Wilk MB. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52(3–4):591–611, <https://doi.org/10.1093/biomet/52.3-4.591>.
32. Posthuma L, van Gils J, Zijp MC, van de Meent D, de Zwart D. 2019. Species sensitivity distributions for use in environmental protection, assessment, and management of aquatic ecosystems for 12 386 chemicals. *Environ Toxicol Chem* 38(4):905–917, PMID: 30675920, <https://doi.org/10.1002/etc.4373>.
33. Hong J, Shaked S, Rosenbaum RK, Jolliet O. 2010. Analytical uncertainty propagation in life cycle inventory and impact assessment: application to an automobile front panel. *Int J Life Cycle Assess* 15(5):499–510, <https://doi.org/10.1007/s11367-010-0175-4>.
34. Venables WN, Ripley BD. 2002. *Modern Applied Statistics with S*. New York, NY: Springer.
35. Rosenbaum RK, Georgiadis S, Fantke P. 2018. Uncertainty management and sensitivity analysis. In: *Life Cycle Assessment*. Hauschild M, Rosenbaum R, Olsen S, eds. Cham, Switzerland: Springer.
36. California Environmental Protection Agency, Office of Environmental Health Hazard Assessment. 2008. *Technical Support Document for the Derivation of Noncancer Reference Exposure Levels. Air Toxic Hot Spots, Risk Assessment Guidelines*. <https://oehha.ca.gov/media/downloads/crn/noncancerstdfinal.pdf> [accessed 17 March 2020].
37. Ring CL, Arnot JA, Bennett DH, Egeghy PP, Fantke P, Huang L, et al. 2019. Consensus modeling of median chemical intake for the U.S. population based on predictions of exposure pathways. *Environ Sci Technol* 53(2):719–732, PMID: 30516957, <https://doi.org/10.1021/acs.est.8b04056>.
38. Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.
39. Aurisano N, Huang L, Milà I, Canals L, Jolliet O, Fantke P. 2021. Chemicals of concern in plastic toys. *Environ Int* 146:106194, PMID: 33115697, <https://doi.org/10.1016/j.envint.2020.106194>.
40. Aurisano N, Fantke P, Huang L, Jolliet O. 2022. Estimating mouthing exposure to chemicals in children's products. *J Expo Sci Environ Epidemiol* 32(1):94–102, PMID: 34188178, <https://doi.org/10.1038/s41370-021-00354-0>.
41. Huang L, Fantke P, Ritscher A, Jolliet O. 2022. Chemicals of concern in building materials: a high-throughput screening. *J Hazard Mater* 424(pt C):127574, PMID: 34799153, <https://doi.org/10.1016/j.jhazmat.2021.127574>.
42. Fantke P, Aurisano N, Bare J, Backhaus T, Bulle C, Chapman PM, et al. 2018. Toward harmonizing ecotoxicity characterization in life cycle impact assessment. *Environ Toxicol Chem* 37(12):2955–2971, PMID: 30178491, <https://doi.org/10.1002/etc.4261>.
43. Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ Sci Technol* 49(14):8804–8814, PMID: 26066997, <https://doi.org/10.1021/acs.est.5b02641>.
44. Kleinstreuer NC, Ceger PC, Allen DG, Strickland J, Chang X, Hamm JT, et al. 2016. A curated database of rodent uterotrophic bioactivity. *Environ Health Perspect* 124(5):556–562, PMID: 26431337, <https://doi.org/10.1289/ehp.1510183>.
45. National Research Council. 2009. *Science and Decisions: Advancing Risk Assessment*. Washington, DC: National Academies Press.
46. Fantke P, Bruinen de Bruin Y, Schlüter U, Connolly A, Bessems J, Kephelopoulou S, et al. 2022. The European exposure science strategy 2020–2030. *Environ Int* 170:107555, PMID: 36244229, <https://doi.org/10.1016/j.envint.2022.107555>.
47. Emará Y, Fantke P, Judson R, Chang X, Pradeep P, Lehmann A, et al. 2021. Integrating endocrine-related health effects into comparative human toxicity characterization. *Sci Total Environ* 762:143874, PMID: 33401053, <https://doi.org/10.1016/j.scitotenv.2020.143874>.
48. Mansouri K, Karmaus AL, Fitzpatrick J, Patlewicz G, Pradeep P, Alberga D, et al. 2021. CATMoS: Collaborative Acute Toxicity Modeling Suite. *Environ Health Perspect* 129(4):47013, PMID: 33929906, <https://doi.org/10.1289/EHP10369>.
49. Fantke P, Illner N. 2019. Goods that are good enough: introducing an absolute sustainability perspective for managing chemicals in consumer products. *Curr Opin Green Sustain Chem* 15:91–97, <https://doi.org/10.1016/j.cogsc.2018.12.001>.
50. Kosnik MB, Hauschild M, Fantke P. 2022. Toward assessing absolute environmental sustainability of chemical pollution. *Environ Sci Technol* 56(8):4776–4787, PMID: 35349278, <https://doi.org/10.1021/acs.est.1c06098>.
51. Fantke P, Cinquemani C, Yaseneva PD, De Mello J, Schwabe H, Ebeling B, et al. 2021. Transition to sustainable chemistry through digitalization. *Chem* 7(11):P2866–P2882, <https://doi.org/10.1016/j.chempr.2021.09.012>.