Probabilistic Reference and 10% Effect Concentrations for Characterizing Inhalation Non-cancer and Developmental/Reproductive Effects for 2,160 Substances

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

Published in:
Environmental Science and Technology

Link to article, DOI:
10.1021/acs.est.4c00207

Publication date:
2024

Document Version
Peer reviewed version

Probabilistic reference and 10% effect concentrations for characterizing inhalation non-cancer and developmental/reproductive effects for 2,160 substances.

Nicolò Aurisano[a], Peter Fantke[a*], Weihsueh A. Chiu[b], Richard Judson[c], Suji Jang[b], Aswani Unnikrishnan[c], Olivier Jolliet[a,d]

[a] Quantitative Sustainability Assessment, Department of Environmental and Resource Engineering, Technical University of Denmark, Bygningstorvet 115, 2800 Kgs. Lyngby, Denmark

[b] Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, USA

[c] National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

[d] Department of Environmental Health Sciences, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA

[*]Corresponding author: Tel.: +45 45254452, e-mail: pefan@dtu.dk

ORCID:
Nicolò Aurisano: https://orcid.org/0000-0003-3651-1307
Peter Fantke: https://orcid.org/0000-0001-7148-6982
Weihsueh A. Chiu: https://orcid.org/0000-0002-7575-2368
Richard Judson: https://orcid.org/0000-0002-2348-9633
Suji Jang: https://orcid.org/0000-0003-1424-2337
Aswani Unnikrishnan: https://orcid.org/0000-0001-9285-8714
Olivier Jolliet: https://orcid.org/0000-0001-6955-4210

Abstract

Chemicals assessment and management frameworks rely on regulatory toxicity values, which are based on points of departure (POD) identified following rigorous dose-response assessments. Yet regulatory PODs and toxicity values for inhalation exposure (i.e., reference concentrations [RfCs]) are available for only 200 chemicals. To address this gap, we applied a workflow to determine surrogate inhalation route PODs, and corresponding toxicity values, where regulatory assessments are lacking. We curated and selected inhalation *in vivo* data from the U.S. EPA's ToxValDB and adjusted reported effect values to chronic human equivalent benchmark concentrations (BMC_h) following the WHO/IPCS framework. Using ToxValDB chemicals with existing PODs associated with regulatory toxicity values, we found that the 25th percentile of a chemical's BMC_h distribution (POD_{p25BMC_h}) could serve as a suitable surrogate for these regulatory PODs ($Q^2 \geq 0.76$, RSE$\leq$0.82 $\log_{10}$ units). We applied this approach to derive POD_{p25BMC_h} for 2,095 substances with general non-cancer toxicity effects and 638 substances with reproductive/developmental toxicity effects, yielding a total coverage of 2,160 substances. From these POD_{p25BMC_h}, we derived probabilistic RfCs and human population effect concentrations. With this work, we have expanded the number of chemicals with toxicity values available thereby enabling a much broader coverage for inhalation risk and impact assessment.

Keywords

- Human toxicity
- Points of departure
- Inhalation Exposure
- ToxValDB
- Chemical substitution
- Life cycle impact assessment

Synopsis

Regulatory toxicity values for inhalation exposure are available for a limited amount of chemicals. This study provides surrogate toxicity values for thousands of substances based on available *in vivo* data.

1. Introduction

Chemical assessment and management frameworks, including life cycle impact assessment (LCIA), and comparative risk screening, evaluate potential risks and toxicological impacts from chemical exposures using chemical-specific points of departure (PODs).[1–4] PODs represent the point on the dose-response curve used for low-dose extrapolation for risk assessment.[5] If the available toxicity data are suitable for dose-response modeling, the statistically-derived benchmark concentration lower confidence limit (BMCL) is modeled and considered as a candidate POD for toxicity value derivation; otherwise, the lowest-observed-adverse-effect concentration (LOAEC) or the no-observed-adverse-effect concentration (NOAEC) are used instead.[6,7] In addition, many frameworks require PODs based on regulatory assessments and thus derived from a comprehensive and systematic dose-response assessment process of available toxicity studies. These include peer-reviewed human health toxicity values from, for example, the U.S. EPA's Provisional Peer Reviewed Toxicity Values (PPRTV), and the Office of Pesticide Programs. Yet, human health assessment relevant data sources currently only provide PODs for a small fraction of the tens of thousands of chemicals used worldwide,[8–11] since conducting such assessments is highly data-, time-, and resource-intensive.[12]

The World Health Organization's International Programme on Chemical Safety (WHO/IPCS) developed a consistent and transparent framework for dose-response assessment that results in the derivation of reference doses (RfDs) and reference concentrations (RfCs) from probabilistically modeled PODs, for both health-based risk assessment as well as comparative risk screening.[13–16] For LCIA purposes, this framework was adopted for deriving human dose-response factors for non-cancer endpoints, using population effect concentrations with an incidence response level $I = 10\%$.[1] Even though it can be applied to derive both RfDs and RfCs, it has mainly been applied to the evaluation of

health risks via the oral route of exposure. Specifically, Chiu et al.[16] derived probabilistic RfDs for 608 substances with assessment-relevant data, and only 1 probabilistic RfC was derived for acrolein.[7] Fantke et al.[1] derived human population effect doses (I = 10%) for 115 organic chemicals, and Aurisano et al.[17] derived probabilistic RfDs and human population effect doses for 10,145 substances. However, no sets of human population effect concentrations for inhalation exposure have been derived yet, mainly due to the much lower data availability of inhalation toxicity studies,[12] as well as the low substance coverage across regulatory sources with RfCs available for $n < 200$ chemicals.

The availability of toxicity values for thousands of chemicals for inhalation exposure is nevertheless crucial, especially for comparing chemicals across exposure routes,[18–20] and for assessing chemicals in a variety of product applications where inhalation often is the predominant exposure route.[21] To address this need, we can take advantage of the increasing availability of experimental animal data housed in databases, such as the U.S. EPA's Toxicity Value Database (ToxValDB), where *in vivo* toxicity data covering inhalation exposure are available for hundreds of chemical substances.[22,23]

In the present study, we propose to adapt the probabilistic risk assessment workflow developed for oral exposures by Aurisano et al.[17] to the derivation of surrogate inhalation PODs, probabilistic RfCs, and human population effect concentrations. We focus on the following four specific objectives:

(i) to compile from ToxValDB a curated dataset of inhalation exposure-response toxicity data covering multiple non-cancer endpoints,

(ii) to develop an approach to derive surrogate inhalation route PODs based on the distribution of available *in vivo* toxicity data in ToxValDB and compare them with available PODs based on regulatory assessments,

(iii) to apply (ii) to derive surrogate inhalation PODs (and their uncertainties) for a wide range of chemicals, separately for general non-cancer effects and reproductive/developmental effects, and

(iv) to use the surrogate inhalation PODs to determine human population effect concentrations at 10% incidence response and probabilistic RfCs using the WHO/IPCS framework and compare the latter against available regulatory RfCs.

We consider two different health effect categories: reproductive/developmental effects and non-reproductive/developmental effects (the latter hereafter referred to as "general non-cancer effects"). This choice is dictated by the large difference between these two categories in levels of severity assigned to the predicted population response levels, i.e., disability-adjusted life year (DALY) estimates associated with different effect types.[1,17,24] The provided set of surrogate inhalation PODs, corresponding RfCs and human population effect concentrations, are suitable for implementation into LCIA, chemical alternatives assessment, and high-throughput risk screening for chemical substitution and prioritization.[1,25,26]

2. Methods

We propose a workflow that aims to derive surrogate inhalation PODs, building on the assumption that for substances for which regulatory toxicity values for inhalation exposure are lacking, probabilistic modeling of available *in vivo* toxicity data might be used for a distribution of chemicals to estimate a POD that most closely mimics statistically a POD that is derived from a regulatory assessment, as done in our previous effort for the oral exposure route.[17] Since in regulatory assessments PODs are usually selected based on a "sensitive" endpoint, surrogate POD would be expected to be at the lower end of the distribution of available toxicity values,[27] following careful data curation where needed.[28] Indeed, for a given

chemical, multiple studies might be available reporting various effect-level types (e.g., BMC, NOAEC), observed critical effects (e.g., mortality, developmental), and tested species (e.g., rabbit, mice), with the consequence of reported effect-level values (i.e., experimental values of toxicity from individual studies) varying over orders of magnitude.[5,16,29] Thus, the numerous challenges in using experimental animal databases need to be overcome by applying methods for data selection and harmonization for human toxicity information,[30,31] similar to those proposed for physico-chemical properties and freshwater ecotoxity information.[32,33]

To compile a harmonized dataset of inhalation exposure-response toxicity data and derive related surrogate POD values, our proposed workflow is composed of six main stages (Fig. 1). The first stage is the curation and selection of the relevant experimental animal toxicity data, and their allocation in one of the two considered health effect categories, i.e., general non-cancer and reproductive/developmental effects (Fig. 1A). Next, we compiled a dataset of reported POD values for inhalation from various regulatory assessments ($\mathrm{POD}_{\mathrm{reg}}$) (Fig. 1B). Third, we investigated the correlation between curated and selected inhalation experimental animal toxicity data and the collected $\mathrm{POD}_{\mathrm{reg}}$ for an overlapping subset of chemicals (Fig. 1C). Based on this analysis, we then systematically determined a surrogate POD for each substance in the two curated datasets (Fig. 1D) and characterized the uncertainty around the determined value (Fig. 1E). Finally, we derived probabilistic RfCs and human population effect concentrations (I = 10%) with related uncertainty using the WHO/IPCS framework (Fig. 1F). The following sections present each of these stages in further detail. For additional information on the applied WHO/IPCS consensus framework, see Supporting Information (SI) Text S1 and Fig. S1.

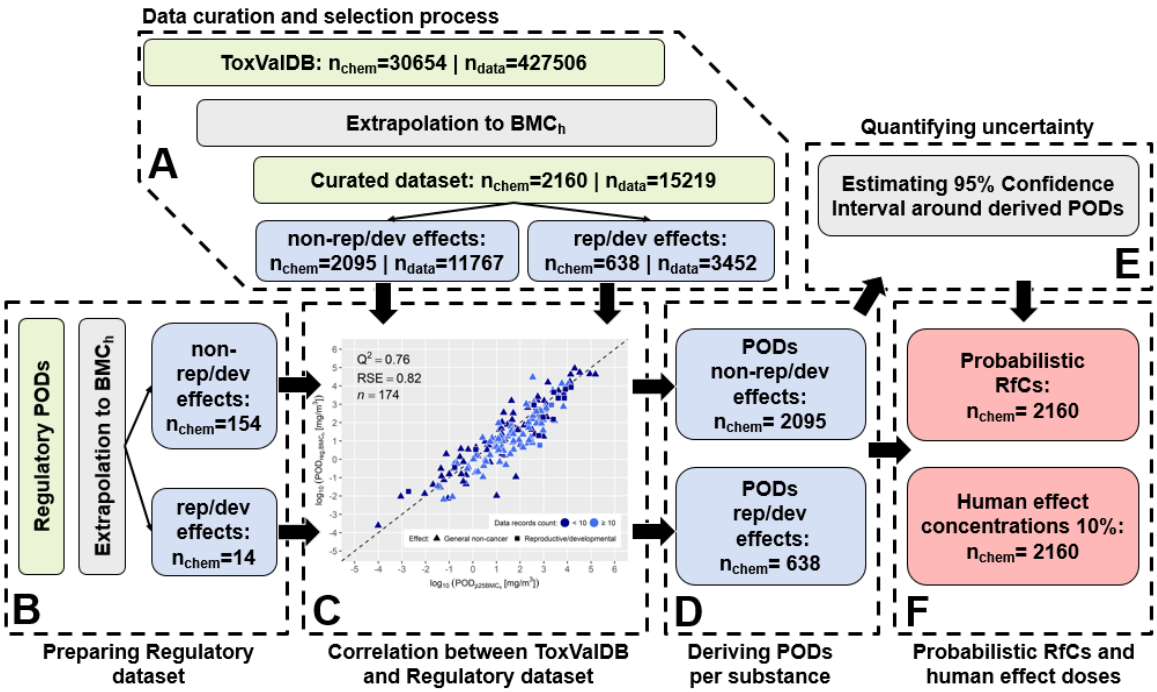


Fig. 1**.** Overview of the proposed workflow: (**A**) data curation and selection applied to the collected *in vivo* data from ToxValDB, (**B**) collection and extrapolation of regulatory PODs, (**C**) analysis of the correlation between ToxValDB and regulatory POD data, (**D**) systematic derivation of surrogate inhalation PODs from the curated datasets, differentiating between general non-cancer (non-reproductive/developmental) and reproductive/developmental effects, (**E**) quantification of the substance-specific uncertainty of the derived PODs, and (**F**) derivation of probabilistic reference concentrations (RfC) and human population effect concentrations at 10% incidence level. The workflow is adapted from Aurisano et al.[17]

2.1. Description of the in vivo input data set

We used ToxValDB as a source for the experimental animal toxicity data (accessible at https://comptox.epa.gov/dashboard).[34,35] The entire ToxValDB was downloaded for subsequent filtering and processing. ToxValDB is a database collecting toxicity data from more than forty publicly available sources,[22] including—among others—ToxRefDB,[36,37] IRIS, PPRTV, ECHA's eChem Portal and the EFSA's Chemical Hazards Database.

2.2. Input data curation and selection

ToxValDB is reporting toxicity data from diverse sources. Such data are often of varying quality, developed for specific applications using different methods, following different guidelines, and available in formats that are not always easily integrated.[32,38,39] For this reason, the collected toxicity data went through a curation and selection process.[14–16] The curation process had three main objectives. Firstly, it aimed to harmonize the reported information, ensuring easy data processing for our study. Secondly, it involved filtering out all records irrelevant to our analysis, specifically focusing on inhalation data. Lastly, it sought to make the reported toxicity animal data directly comparable across different species and study types. The reported effect value derived for each record was extrapolated to a chronic human equivalent benchmark concentration (BMC_h) with a consistent unit expressed in mg/m^3. The extrapolations covered LOAEC-to-NOAEC, NOAEC-to-BMC, BMCL-to-BMC, exposure duration extrapolation, and the application of a dosimetric adjustment factor. For details on the data curation and selection as well as an overview of the extrapolation factors applied, see SI Text S2 and Tables S1-S2. After the curation and selection process, the curated data were split into two distinct datasets covering general non-cancer and reproductive/developmental effects based on each record study type and reported critical effects (Fig. 1A).

2.3. Regulatory data

To build a regulatory dataset for inhalation exposure, we used as a starting point the work conducted by Wignall et al.,[12,40] collecting peer-reviewed toxicity values reported in various public sources, such as IRIS. In addition, the collected toxicity values were cross-checked with the November 2019 release of the U.S. EPA RSL and incorporated new substances with available PODs.[41] In our study, a POD_{reg} is defined as an inhalation exposure route NOAEC, LOAEC, or BMCL associated with a reported RfC collected from one of the

198 above-mentioned data sources. As done for the *in vivo* input data set, POD_{reg} values were
199 extrapolated to chronic human equivalent benchmark concentrations (POD_{reg,BMC_h}) (Fig. 1B),
200 see SI Text S2 and Tables S1-S2.
201
202 **2.4. Approach for deriving surrogate inhalation PODs**
203 To derive surrogate inhalation PODs, we followed and tested the hypothesis that, for
204 each chemical, the lower end of the effect values distribution available in ToxValDB is
205 considered as a suitable proxy for POD_{reg}.[27] Thus, for chemicals for which both POD_{reg,BMC_h}
206 and *in vivo* data were available, we assumed a lognormal distribution across BMC_h and
207 separately derived the 5th %-ile, 15th %-ile, 25th %-ile and 35th %-ile of the fitted lognormal
208 distribution. To test the appropriateness of the selected percentile for inhalation toxicity data,
209 these different percentiles were compared against the respective available POD_{reg,BMC_h} showing
210 that the 25th %-ile is the percentile with the lowest bias when regressed against the regulatory
211 values. Since the intercept was not significantly different from zero, we tested a regression
212 forcing the intercept to zero, yielding a slope of 0.99 (95th CI, 0.94 – 1.05), not significantly
213 different from 1 (Table S3). The 25th %-ile of the human benchmark concentrations was
214 therefore directly selected as the inhalation POD (POD_{p25BMC_h}). This is consistent with our
215 previous study on oral exposure,[17] that also identified the 25th %-ile of the benchmark dose as
216 the most suitable for estimating a surrogate oral POD.
217 The two function moments used for fitting the lognormal distribution to BMC_h values
218 are mu (μ) and sigma (σ), where μ represents the log-scale population median, and σ is the
219 standard deviation of the available effect values for a substance.[42] μ was calculated from the
220 available BMC_h for all substances. Whereas, for σ, due to the highly unstable estimates of σ
221 for chemicals with a limited amount of records available (Fig. S2), an average-shaped
222 distribution was applied instead of relying on the few available effect values to avoid bias

introduced by too few data points. Thus, we differentiate between substances with ≥10 records available (data-rich chemicals) and <10 records available (data-poor chemicals), after extrapolating all data to chronic values. For fitting the lognormal distribution, σ was calculated from the available BMC_h only for data-rich chemicals, whereas for data-poor chemicals, we applied a fixed standard deviation (σ_{fixed}). The σ_{fixed} is derived from the average across σ of data-rich chemicals. This is consistent with our previous study on oral exposure.[17] Where despite observing considerable variability across effect values available for substances with the same number of records, largely attributed to disparities in underlying data, our analysis revealed a consistent trend: as the number of records increased, this variability steadily diminished. We identified 10 records as a reliable and pragmatic cutoff point.

2.5. Deriving surrogate inhalation PODs across ToxValDB substances without regulatory values

After confirming that the 25th %-ile of the fitted lognormal distribution is suitable to derive surrogate inhalation route PODs, we systematically estimated μ and σ for each substance from the available records and then derived related POD_{p25BMC_h} (Fig. 1D). This was done separately for the two categories of effects. For substances with curated toxicity records available in both categories, two distinct POD_{p25BMC_h} values were derived.

2.6. Quantifying uncertainty around the derived points of departure

We characterized the uncertainty around the derived POD_{p25BMC_h} from the residual standard error (RSE) of the comparison carried out between POD_{reg} and POD_{p25BMC_h} (Fig. 1E). This uncertainty is expressed as the squared geometric standard deviation ($GSD^2_{p25\rightarrow reg}$). $GSD^2_{p25\rightarrow reg}$ describes the spread of data around their geometric mean, and more specifically

indicates that 95% of the data fall within the range of $\mathrm{POD}_{p25\mathrm{BMC_h}}/\mathrm{GSD}^2_{p25\rightarrow\mathrm{reg}}$ and $\mathrm{POD}_{p25\mathrm{BMC_h}} \times \mathrm{GSD}^2_{p25\rightarrow\mathrm{reg}}$.[43–46] For example, a $\mathrm{GSD}^2_{p25\rightarrow\mathrm{reg}} = 10$ indicates that the 95% confidence interval of $\mathrm{POD}_{p25\mathrm{BMC_h}}$ span over two orders of magnitude.

2.7. Deriving probabilistic reference and human effect concentrations

For each of the derived $\mathrm{POD}_{p25\mathrm{BMC_h}}$ we calculated probabilistic RfCs and human population effect concentrations (at incidence level I = 10%) (Fig. 1F).

We implemented the approximate approach by Chiu et al.[16] for the calculation of the probabilistic RfCs. These were derived from the lower 95% confidence bound of $\mathrm{HC_M}^{1\%}$, i.e., the daily human concentration at which, with 95% confidence, no more than 1% of the population shows a level of effect M corresponding to the effect level type reported in the database and the type of endpoint (e.g., continuous, quantal deterministic, or stochastic deterministic). For each chemical, $\mathrm{HC_M}^{1\%}$ was calculated from $\mathrm{POD}_{p25\mathrm{BMC_h}}$ by dividing it by an extrapolation factor of $P50 = 9.7$ to account for variability in sensitivity between the median and the 1st %-ile of human population response.[14] The lower 95% confidence bound of $\mathrm{HC_M}^{1\%}$ was derived by combining probabilistically $\mathrm{GSD}^2_{p25\rightarrow\mathrm{reg}}$ and the uncertainty distribution (i.e., $P95/P50 = \mathrm{GSD_h}^{1.65} = 4.3$) assigned to the human variability at 1st %-ile.[14] Note that $\mathrm{GSD}^{1.65} = (\mathrm{GSD}^2)^{\frac{1.65}{2}}$, indicating a one-sided (i.e., lower) confidence interval range. We then compared the derived lower 95% confidence bound of $\mathrm{HC_M}^{1\%}$ against the related regulatory RfC (if available) to investigate the potential influence of the database uncertainty factor (UF_d). UF_d is commonly implemented when deriving regulatory RfCs but is not directly included in the WHO/IPCS framework.[16] Nevertheless, in our previous study focusing on oral exposure,[17] an approach was developed for including such uncertainty factors to account for data gaps as a function of records availability. We implemented and

tested the same approach in the present study to understand whether the derived toxicity values are consistent with regulatory RfCs and identify potential biases.

For LCIA purposes, recent updates of the globally recommended approach for deriving human dose-response factors for non-cancer endpoints proposed using the human population effect concentration with an incidence response level I = 10%.[1] We derived $HC_M^{10\%}$ from the provided POD_{p25BMC_h}, accounting for the human variability between 50% and 10% incidence level by dividing the POD_{p25BMC_h} by the best estimate factor of $P50 = 3.49$.[14] $HC_M^{10\%}$ related uncertainty was calculated by combining probabilistically $GSD^2_{p25\rightarrow reg}$ of POD_{p25BMC_h} and the uncertainty distribution assigned to the human variability at 10th %-ile, i.e., $P97.5/P50 = 2.67$.[14]

Finally, the derived RfCs and $HC_M^{10\%}$s were compared against the results of our previous study on oral toxicity to investigate potential trends across exposure routes.[17] In addition, we matched our results with exposure estimates from the Systematic Empirical Evaluation of Models (SEEM) meta-model.[47] In this analysis, we aimed to identify the fraction of assessed substances with population median chemical intake rates above our derived probabilistic RfCs to put the obtained results into perspective. Indeed, the identified substances will deserve further scrutiny, since for these exposure best estimates are higher than derived probabilistic RfCs, highlighting a high potential risk. For these comparisons, RfCs and $HC_M^{10\%}$ were converted into a consistent unit of mg/kg-d, assuming an average individual human breathing rate of 13 m^3/d and body weight of 70 kg.

2.8. Data analysis

The gathered toxicity data from ToxValDB were processed using the open source statistical software R version 3.6.1,[48] and the package "ggplot2" was used to generate all

results figures.[49] The R code used in this study for deriving PODs from the provided selected and harmonized datasets is available in SI Text S3.

3. Results

3.1. Curated toxicity test datasets

The downloaded version of ToxValDB listed 427,506 records for more than 30,000 chemicals and reported a wide range of toxicity information. The resulting curated dataset compiled inhalation toxicity information for 2,160 substances covered by 15,219 records. We split this curated dataset into two distinct datasets covering two health effect categories, i.e., 2,095 substances (11,767 records) for general non-cancer and 638 substances (3,452 records) for reproductive/developmental effects. Records were available in both datasets for 573 substances. The curated datasets are provided in the SI, separately for general non-cancer effects (Table S4) and reproductive/developmental effects (Table S5).

Fig. S3 summarizes the statistics of the two curated datasets. Fig. S3A-B presents the distribution of the extrapolated effects values (BMC_h) across all records, differentiating between underlying effect-level and study types information. NOAEC is the most common effect-level type in both datasets (~75%), followed by LOAEC (~24%) and BMCL (~1%). For the study types distribution across records in the general non-cancer effects dataset (Fig. S3A), 33%, 58%, and 8% of the records were reported as chronic, subchronic, and subacute, respectively.

Fig. S3C-D gives an overview of the number of curated records available per substance in the two datasets, highlighting the limited number of records available for most substances. For example, only one or two records are available for around half of the covered substances, and for both datasets, only 15% of all substances are considered as data-rich chemicals. Concerning the tested species, the majority of records report rat followed by

mouse (Fig. S3E-F). The statistics of the curated datasets are in line with data for oral exposure,[17] and with other studies using the same database to develop QSAR and new approach methodologies (NAMs) models.[5,6]

Fig. 2A-B visualizes the extrapolated effect values (BMC_h) for all the records in the two datasets, differentiating between originally reported effect-level types. Across the records in the general non-cancer effects, BMC_h values range from 2.5×10^{-4} to 3.7×10^{7} mg/m^3 with a median value of 713 mg/m^3 (Fig. 2A), while in the reproductive/developmental effects dataset, they range from 3.5×10^{-3} to 3.4×10^{6} mg/m^3 with a median value of 7875 mg/m^3 (Fig. 2B). The BMC_h values across the records available for the same substance can span over several orders of magnitude. In general, this variability might be related to different factors, such as different critical effects studied or species tested in various environmental conditions (i.e., biological variability), as well as systematic errors, including errors in the measurements, differences in experimental protocols, or measurement tools.[5,50,51] In addition, collected POD_{reg,BMC_h} are represented as black triangles for the substances for which a regulatory RfC was available.

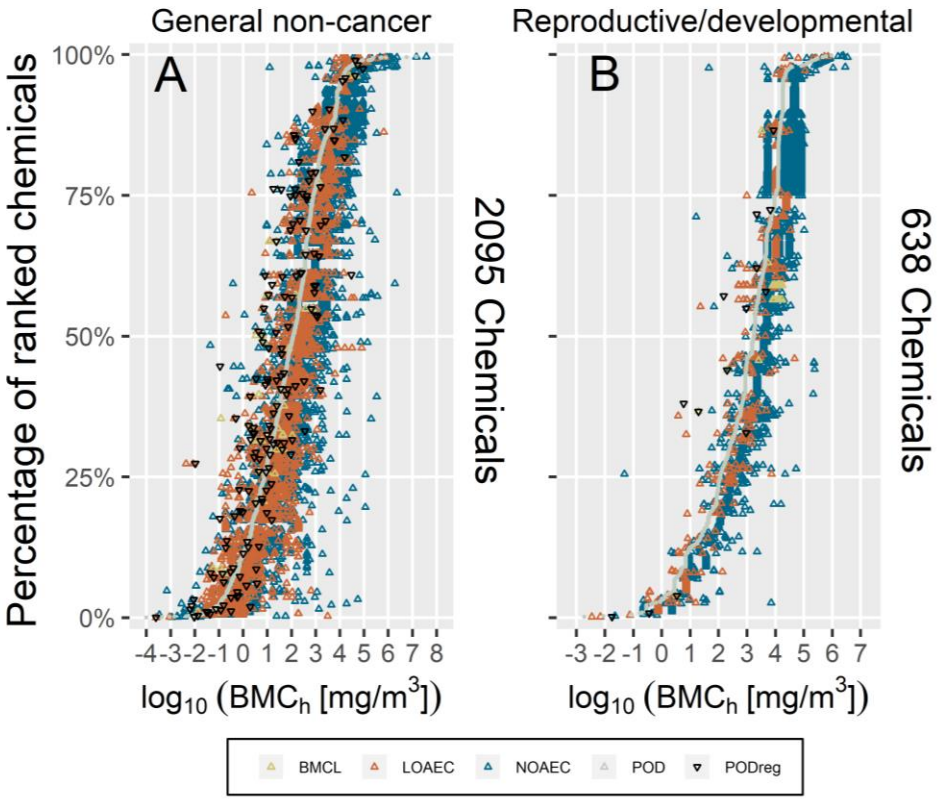


Fig. 2**.** Inhalation effect values adjusted to chronic human equivalent benchmark concentrations (BMC_h) for all the records in the general non-cancer effects (**A**) and reproductive/developmental effects (**B**) dataset, together with the corresponding POD_{reg} (black ▽, when available) and derived PODs (POD_{p25BMC_h}, grey data points). Chemicals are ranked by derived POD_{p25BMC_h}.

3.2. Points of departure and comparison with regulatory toxicity values

We derived surrogate inhalation PODs for all the substances for which toxicity information were available in the curated datasets as the 25th %-ile of the fitted lognormal distribution to the available records per substance (POD_{p25BMC_h}). To fit the lognormal distribution for data-rich chemicals (≥10 records), we directly used the available effect values (BMC_h) to derive a chemical-specific standard deviation, assuming that the available records are sufficient to represent and cover different potential adverse effects. In contrast, we derived an average standard deviation across data-rich chemicals of $\log_{10}\sigma_{fixed} = 0.6$ for both

general non-cancer and reproductive/developmental effects (Fig. S2). We then applied σ_{fixed} to all data-poor chemicals (<10 records) for calculating $\text{POD}_{p25\text{BMC}_\text{h}}$.

We systematically derived surrogate $\text{POD}_{p25\text{BMC}_\text{h}}$ for 2,095 substances for general non-cancer effects and 638 substances for reproductive/developmental effects, yielding a total substance coverage of 2,160. For 573 substances, we derived two distinct $\text{POD}_{p25\text{BMC}_\text{h}}$ as toxicity values were available for both health effect categories. The derived $\text{POD}_{p25\text{BMC}_\text{h}}$ are presented in Fig. 2, ranging from 9.8×10^{-5} to 1.5×10^{7} mg/m^3 for general non-cancer effects, with a median $\text{POD}_{p25\text{BMC}_\text{h}}$ value of 117 mg/m^3. For reproductive/developmental effects, $\text{POD}_{p25\text{BMC}_\text{h}}$ are on average more than one order of magnitude higher, ranging from 1.9×10^{-3} to 9.4×10^{5} mg/m^3. Table S6 provides all derived PODs.

Fig. 3 compares the derived $\text{POD}_{p25\text{BMC}_\text{h}}$ against the respective available $\text{POD}_{\text{reg,BMC}_\text{h}}$ for both studied effects. The comparison was carried out for a total of $n = 174$ substances with available regulatory inhalation data, i.e., $n = 160$ substances for general non-cancer (Fig. S4A) and $n = 14$ for reproductive/developmental effects (Fig. S4B). The estimated $\text{POD}_{p25\text{BMC}_\text{h}}$ values correlate well with the derived $\text{POD}_{p25\text{BMC}_\text{h}}$, with a coefficient of "goodness of prediction" of $Q^2 = 1 - PRESS/TSS = 0.76$ and a residual standard error on the log of RSE $= 0.82$ evaluated on log-scale for the 1:1 line. PRESS is the Predictive Error Sum of Squares, which is the sum of the squares of the differences (residuals) between the predicted and regulatory values, and TSS is the Total Sum of Square.[52] To evaluate the choice of selecting the 25$^{\text{th}}$ %-ile of the fitted lognormal distribution as a surrogate for regulatory data, we analyzed the correlation of $\text{POD}_{\text{reg,BMC}_\text{h}}$ against other three additional percentiles, i.e., $\text{POD}_{p05\text{BMD}_\text{h}}$, $\text{POD}_{p15\text{BMD}_\text{h}}$, $\text{POD}_{p35\text{BMD}_\text{h}}$ (Fig. S5). This analysis showed that the 25$^{\text{th}}$ %-ile is the percentile with the lowest bias when regressed against the regulatory values. Since the intercept was not significantly different from zero, we tested a regression forcing the intercept to zero, yielding a slope of 0.99 (95$^{\text{th}}$ CI, 0.94 – 1.05), not significantly different

from 1 (Table S3). This supports our choice of the 25th %-ile, which is consistent with the best-suited percentile identified for oral exposure.[17] The collected POD_{reg,BMC_h} are summarized in Table S7.

For around 60% ($n = 104$) of the substances the derived POD_{p25BMC_h} values are slightly higher than the respective available POD_{reg,BMC_h}. This suggests that the provided POD_{p25BMC_h} values might be slightly less conservative than regulatory ones based on the best-fitted %-ile, which is reflected in the uncertainty factor derived from RSE and applied when deriving probabilistic reference and human effect concentrations.

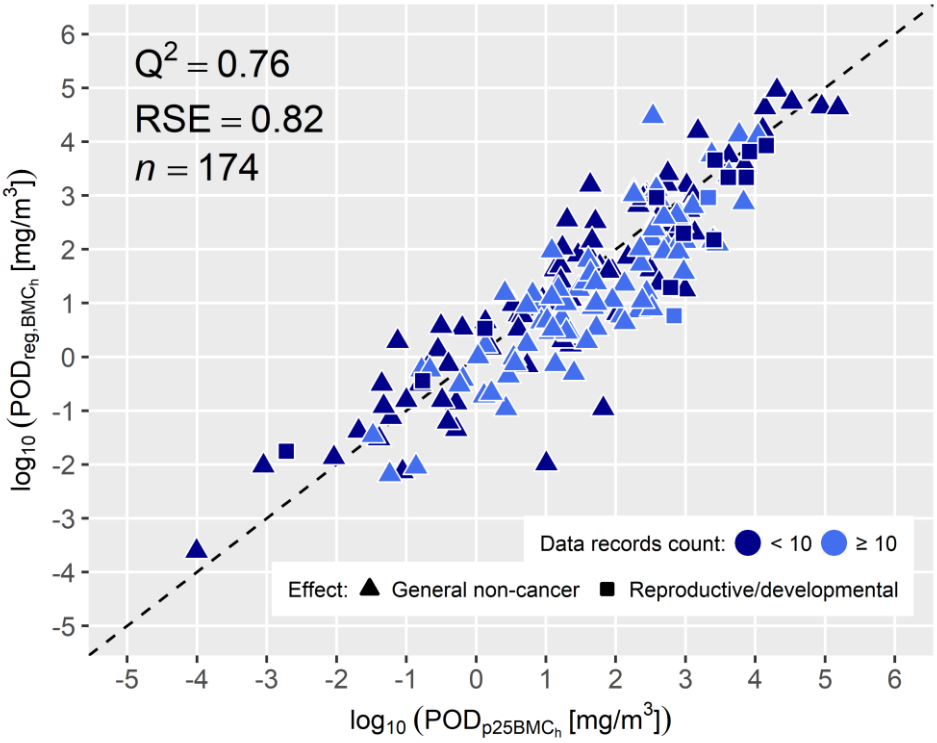


Fig. 3. Comparison between estimated POD_{p25BMC_h} and POD_{reg,BMC_h} for general non-cancer (▲) and for reproductive/developmental effects (■), differentiating between data-rich (light blue, ≥10 records) and data-poor chemicals (dark blue, <10 records). Coefficient of "goodness of prediction" (Q^2) and residual standard error (RSE) are evaluated on log-scale for the 1:1 line (black dashed line).

The uncertainty factor of $GSD^2_{p25\rightarrow reg} = 10^{2.02x0.82} = 45$ is derived from this comparison with regulatory values, to reflect the use of POD_{p25BMD} as a suitable approximation of POD_{reg}. The limited amount of chemicals considered in the comparison against regulatory values for reproductive/developmental effects precluded the characterization of an effect-specific uncertainty; hence, the same uncertainty as for general non-cancer effects is used by default. Fig. S6 presents the distributions of the derived POD_{p25BMD_h} together with their characterized 95% CI.

3.3. Probabilistic reference concentrations and human population effect concentrations

From the provided POD_{p25BMC_h}, we derived probabilistic RfCs and human population effect concentrations ($HC_M^{10\%}$), following the WHO/IPCS framework.[14] Since WHO/IPCS focus on endpoint-specific uncertainties and RfCs, an additional database uncertainty factor (UF_d) needed to be included when deriving probabilistic RfCs that are comparable to and consistent with regulatory RfCs. UF_d accounts for data gaps and is typically equal to 1, 3, and 10 as a function of the data coverage for different endpoints.[53] Since access to the underlying data of each chemical is limited in the ToxValDB, chemical-specific data availability was used as a surrogate to deterministically estimate additional UF_d following the approach of Aurisano et al.[17]: the lower 95% confidence bound of $HC_M^{1\%}$ is 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 RfC value is thus equal to the lower 95% confidence bound of $HC_M^{1\%}$. The derived probabilistic RfCs show a good correlation with the regulatory RfCs with a $Q^2 = 0.59$ and $RSE = 1.11$ evaluated on log-scale for the 1:1 line (Fig. S7B). In the comparison, out of n = 174 substances with available

regulatory inhalation data, $UF_d = 10$ was applied to $n = 42$, $UF_d = 3$ was applied to $n = 64$, and $UF_d = 1$ was applied to $n = 68$ data records, respectively. In contrast, neglecting UF_d would lead to a systematic overestimation of the RfCs (Fig. S7A, $Q^2 = 0.54$, RSE $= 1.18$). Following the UF_d complemented approach, probabilistic RfCs for $n = 2{,}169$ substances were derived.

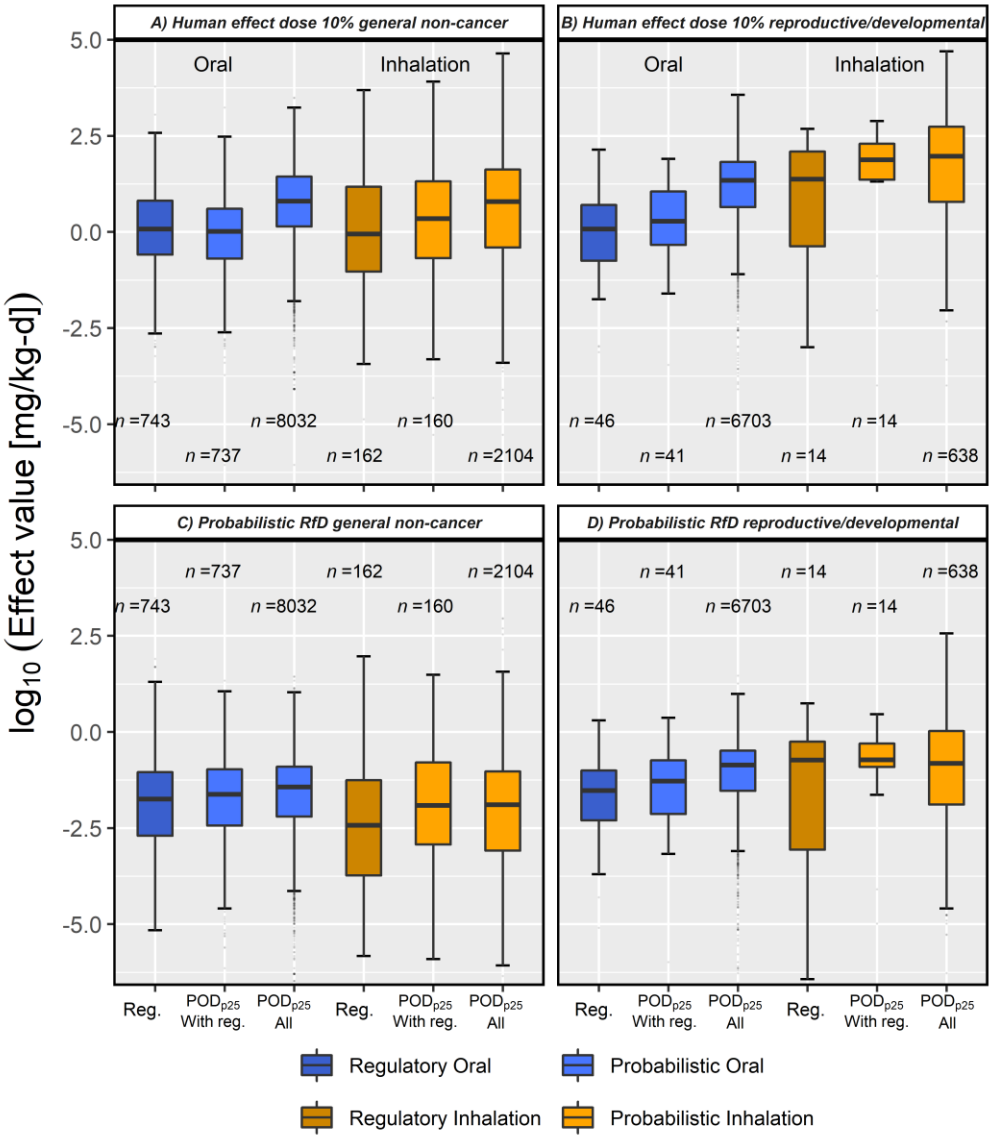
We also derived best estimates of $HC_M^{10\%} = HC_M^{50\%}/3.49 = POD_{p25BMC_h}/3.49$ with their uncertainties.[1] The associated uncertainty characterized by combining probabilistically $GSD^2_{p25\rightarrow reg}$ and the uncertainty distribution assigned to the human variability at 10th %-ile is equal to $GSD^2_{HC_M^{10\%}} = 51$, indicating that the 95% confidence bound of $HC_M^{10\%}$ values on a log-scale is +/- 1.7 $\log_{10}$ mg/m^3 (Fig. S8). Table S6 provides the derived probabilistic RfCs and $HC_M^{10\%}$s with related uncertainties.

3.4. Comparison of toxicity value ranges across health effect categories and exposure routes

Fig. 4 summarizes the derived toxicity values for inhalation and compares their ranges against results for oral exposure provided by Aurisano et al.[17] Fig. 4A-B presents ranges of inhalation $HC_M^{10\%}$ and oral $HD_M^{10\%}$, while Fig. 4C-D presents RfCs and RfDs, separately for the two health effect categories. For each combination (e.g., inhalation $HC_M^{10\%}$ for general non-cancer effects), regulatory values are presented first (darker color), followed by probabilistic values for the same chemicals for which regulatory assessments were available, and finally by the probabilistic values for all covered substances. Regulatory-based $HC_M^{10\%}$ and $HD_M^{10\%}$ are also estimated from POD_{reg} following the WHO/IPCS framework. For these comparisons, RfCs and $HC_M^{10\%}$ were converted into a consistent unit of mg/kg-d, assuming an average individual human breathing rate of 13 m^3/d and body weight of 70 kg.

Fig. 4 confirms that considering chemicals for which regulatory toxicity values were available, the ranges of derived probabilistic values are well in line with regulatory values across different toxicity values, exposure routes, and effects considered. On the other hand, when considering all chemicals, the median of the probabilistic toxicity values is higher than the available regulatory values in the majority of the cases. This trend is linked to our probabilistic results covering thousands of substances while the regulatory values only cover a few hundred substances, and suggests that regulatory values tend to be selected among the most toxic substances.

No discernible trends in ranges across different exposure routes were noted. As an additional step, a more detailed investigation was undertaken to explore potential trends and correlations at the chemical-specific level for substances with available toxicity data for both oral and inhalation exposure. However, this analysis further supports that there are no clear correlations between oral and inhalation toxicity for both general non-cancer and for reproductive/developmental effects (Fig. S9). These low correlations across diverse exposure routes could stem from several factors, such as the consideration of data points coming from a wide range of studies with different settings and examining different critical effects.



456

457 Fig. 4**.** Comparison between oral toxicity values covering probabilistic reference doses (RfDs)
458 and human effect doses (I = 10%, $HC_M^{10\%}$) and the derived probabilistic reference
459 concentrations (RfCs) and human effect concentrations (I = 10%, $HC_M^{10\%}$) for general non-
460 cancer effects (**A** and **C**) and reproductive/developmental effects dataset (**B** and **D**).[17] Note
461 that probabilistic RfCs and $HC_M^{10\%}$ of this study were converted to doses with a consistent
462 unit of mg/kg-d, assuming an average individual human breathing rate of 13 m^3/d and body
463 weight of 70 kg. *n* represents the number of chemicals covered. Reg.: regulatory values;
464 POD_{p25} With reg.: probabilistic values for the same chemicals for which regulatory
465 assessments were available; POD_{p25} All: probabilistic values for all covered substances.

Finally, we compared the derived probabilistic RfCs with population median chemical intake rates estimated via the SEEM meta-model, available for around half of the considered substances.[47] We identified 33 substances for which exposure best estimates are higher than derived probabilistic RfCs, highlighting a high potential risk (Fig. 5). These substances include, for example, insecticides and biocides such as parathion (CAS: 56-38-2) and acrolein (CAS: 107-02-8) as well as substances with various industrial applications such as 1,6-diisocyanatohexane (CAS: 822-06-0). In contrast to our previous study, where only three substances were identified as potentially risky via oral exposure, a greater number of substances were flagged as potentially risky via inhalation exposure in our present study, despite having five times less data for inhalation.[17] When considering the upper 95% confidence bound of the SEEM estimates (grey error bars in Fig. 5), median intake rates are 100 times higher than doses calculated from probabilistic RfCs for around 5% of the substances for which SEEM intake rates are available, that is substances that should be prioritized for further analysis.

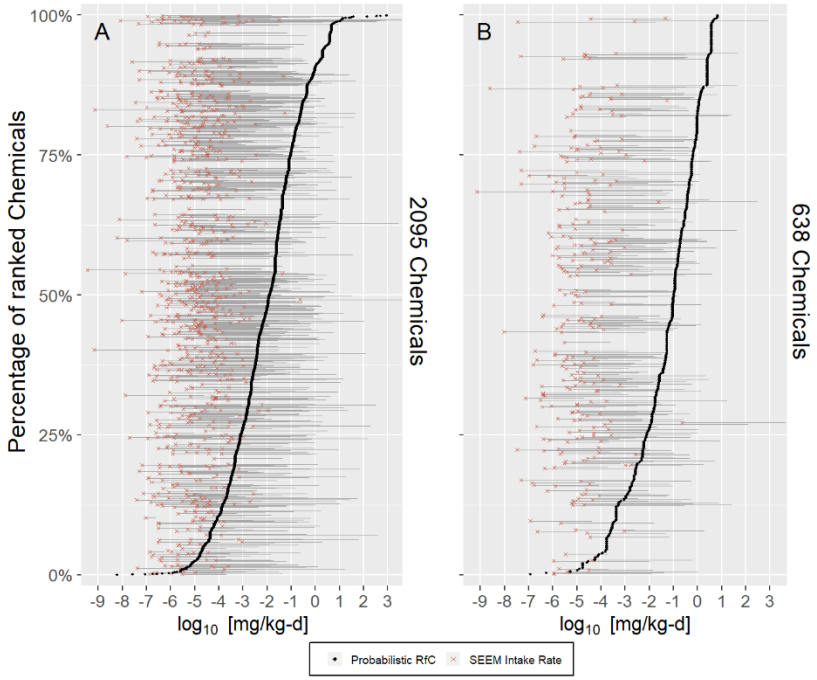


Fig. 5**.** Comparison between probabilistic reference concentrations (RfC) and population median chemical intake rates, differentiating between (**A**) general non-cancer effects and (**B**) reproductive/developmental effects. The upper 95% confidence bound of the SEEM Intake rates (error bars) reflects uncertainty around the population median intake rate and does not reflect population variability. Substances are ranked in increasing order based on the probabilistic RfCs. Probabilistic RfCs were converted into a consistent unit of mg/kg-d, assuming a breathing rate of 13m^3/d and a body weight of 70 kg.

4. Discussion

4.1. Applicability of the derived toxicity values

This study expanded by a factor 13 the coverage of chemicals for which inhalation toxicity values can be derived for general non-cancer effects, and by a factor 45 for reproductive/developmental effects. Combined with our previous effort focused on oral exposure, the presented approach provides a basis for consistently assessing toxicity effects across these two exposure routes for thousands of chemicals in various impact assessment and risk screening contexts.

The provided $HC_M^{10\%}$ can be implemented in LCIA to derive human toxicity effect factors with direct application in USEtox. USEtox is the UNEP/SETAC scientific consensus model for human toxicity and ecotoxicity characterization in life cycle impact assessment and other comparative assessments,[54,55] and aims to improve the understanding and management of chemicals by quantifying exposure, risks, and impacts of chemicals in products (e.g., personal care, toys, building materials) and in the environment.[1,56] USEtox applications include life cycle assessment, chemical footprinting, risk screening, safe and sustainable-by-design (SSbD) and chemical substitution, to inform public and private stakeholders. In the current version of USEtox, human toxicity effect factors covering non-cancer toxicity were

only available for less than 500 chemicals, of which only one-tenth is derived from inhalation toxicity data. The provided $HC_M^{10\%}$s will increase the chemical coverage for inhalation by a factor of forty. In addition, by providing $HC_M^{10\%}$s specific to general non-cancer effects and reproductive/developmental effects, these will be able to reflect the difference in severity when evaluating DALY related to chemical exposure (i.e., 2.4 DALY/incidence for general non-cancer effects vs. 44.1 DALY/incidence for reproductive/developmental effects).[1,24]

The provided probabilistic RfCs find direct application to support high-throughput risk screening studies, where hundreds (if not thousands) of chemicals are assessed in terms of multi-pathway exposure and related effects on humans. Thus the availability of toxicity information is a key factor when, e.g., evaluating exposures and identifying chemicals of concern and potential alternatives to harmful chemicals present in consumer products.[18–20]

Finally, by estimating surrogate inhalation PODs and deriving corresponding toxicity values also for chemicals with a limited amount of toxicity data available, our results support the work of health assessors at multiple levels, including the cases of chemicals of potential concern not yet tested or reviewed.[12] While our results are primarily applicable at screening level and cannot substitute the rigorous assessments of chemicals potentially of concern, they constitute a useful dataset to train *in-silico* approaches beyond the restricted availability of regulatory values.

4.2. Limitations of the proposed workflow

The presented workflow also comes with limitations. First, the provided PODs (and related toxicity values) are based for 85% of the covered substances on less than ten curated records. For data-poor chemicals, there is the possibility of missing critical effects not covered by the considered studies and thus underestimating the toxicity of specific substances. To address this issue when deriving PODs for these substances, we fitted a lognormal distribution with a predefined average shape with a fixed standard deviation.

Nevertheless, fitting a chemical-specific distribution based on a set of experiments to derive $\mathrm{POD}_{p25\mathrm{BMC_h}}$ would still be preferred and more accurate. In our approach, data richness is nevertheless considered, but in a simplified way when deriving probabilistic RfCs, where different UF_d are applied deterministically to the lower 95% confidence bound of $\mathrm{HC_M}^{1\%}$ based on the data availability. Alternative ways to derive UF_d should be explored in the future to account not only for the number but also the type of data available.

In addition, we acknowledge that ToxValDB is reporting toxicity data from diverse sources and as a consequence, such data are often of varying quality and relevance. Except for addressing duration extrapolation (e.g., subchronic to chronic), there is no filtering implemented or prioritization for records with specific quality or reliability (e.g., coming from specific data sources). This is an intrinsic limitation of the database (e.g., NOAELs and NOELs often do not include specific severity/endpoint information), so in our approach, all records are selected and harmonized (e.g., filtering out all records not covering the inhalation route) are treated as equally relevant for further processing. While prioritizing records with higher reliability would lead on the one hand to a higher quality dataset, it would on the other hand lead to a lower chemical coverage as a trade-off; hence, both aspects generally will have to be balanced. Additionally, this concern is somewhat ameliorated by the calibration of our surrogate PODs to authoritative values, so on average we have demonstrated our approach to be unbiased.

The characterized uncertainty for each POD, expressed as $\mathrm{GSD}^2_{p25\rightarrow\mathrm{reg}}$, is limited to the uncertainty around the derived POD and directly reflects the use of $\mathrm{POD}_{p25\mathrm{BMC_h}}$ as a suitable approximation of regulatory values ($\mathrm{POD}_{\mathrm{reg,BMC_h}}$). The limited availability of reproductive/developmental effects data precluded the possibility of deriving an effect-specific $\mathrm{GSD}^2_{p25\rightarrow\mathrm{reg}}$ instead of a generic uncertainty applied to $\mathrm{POD}_{p25\mathrm{BMC_h}}$ for both effects. The same is valid for the uncertainty around the derived reproductive/developmental $\mathrm{HC_M}^{10\%}$s

as well as the uncertainty used to define the lower 95% confidence bound of $HC_M^{1\%}$. Compared to the uncertainties for surrogate PODs derived in our previous study,[17] $GSD^2_{p25\rightarrow reg}$ is higher for inhalation by a factor 3 and up to a factor 6.

Finally, we acknowledge that in our workflow, there is an intrinsic limitation related to predicting a toxicity value from *in vivo* data. More specifically, even if starting from the same underlying toxicity dataset, risk estimates can vary across regulatory settings despite the rigorous scientific judgment involved.[12,59]

4.3. Future research needs

Future research should focus on further increasing the exposure route coverage. Even though the toxicity data availability and related chemical coverage will be lower for other exposure routes, such as dermal, route-specific toxicity data are key for assessing chemicals in specific product applications.

Similarly, in our work, we differentiated between reproductive/developmental effects and general non-cancer effects due to the difference in both the exposure windows involved and the severity of these two disease categories to affect human lifetime loss.[1,24] Another reason for considering only these two effect categories in this study is that for the majority of the globally marketed chemicals, only very few toxicity data points are available—hence, expanding the scope to include more effect categories would have inevitably resulted in a reduction in the number of chemicals covered within each category. This would have reduced precision into the overall comparison by skewing the representation of chemicals across different categories of effects. Future work should further increase this differentiation and provide more critical effect-specific PODs (and related $HC_M^{10\%}$ and probabilistic RfCs). Highly relevant critical effects include, for example, endocrine disruption.[60]

Finally, given the large number of new and existing substances requiring assessment and management, there is a pressing need for cost-effective and rapid non-animal

alternatives.[61] In answer to this, the curated dataset compiling inhalation toxicity information provided in this study can be used in future research for training *in silico*, machine learning-based methods (e.g., random forest algorithms) to construct QSAR models for predicting PODs for substances lacking *in vivo* data.[12,62] This would cover an even broader range of chemical substances.

Supporting Information

- Details covering the data curation and selection process, introduction to the WHO/IPCS consensus framework, R code for deriving points of departure (PODs) from the curated datasets, and additional visualizations of results, including comparison between probabilistic RfCs and SEEM Intake rates (PDF)
- Curated and selected toxicity records from ToxValDB, Derived PODs, probabilistic RfCs, human population effect concentrations, and collected regulatory toxicity records (Excel)

Acknowledgments

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 U.N. Environment's Strategic Approach to International Chemicals Management (SAICM)' (GEF project ID 9771, grant no. S1-32GFL-000632), by the 'Safe and Efficient Chemistry by Design (SafeChem)' project funded by the Swedish Foundation for Strategic Environmental Research (grant no. DIA 2018/11), and by the PARC project (Grant No. 101057014) funded under the European Union's Horizon Europe Research and Innovation program.

Conflict of Interest

The authors declare no conflict of interest.

Disclaimer

Dr. Jolliet 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. 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.

References

(1) Fantke, P.; Chiu, W. A.; Aylward, L.; Judson, R.; Huang, L.; Jang, S.; Gouin, T.; Rhomberg, L.; Aurisano, N.; McKone, T.; Jolliet, O. Exposure and Toxicity Characterization of Chemical Emissions and Chemicals in Products: Global Recommendations and Implementation in USEtox. *Int. J. Life Cycle Assess.* **2021**, *26* (5), 899–915. https://doi.org/10.1007/s11367-021-01889-y.

(2) Fantke, P.; Huang, L.; Overcash, M.; Griffing, E.; Jolliet, O. Life Cycle Based Alternatives Assessment (LCAA) for Chemical Substitution. *Green Chem.* **2020**, *22* (18), 6008–6024. https://doi.org/10.1039/d0gc01544j.

(3) Pham, L. L.; Sheffield, T. Y.; Pradeep, P.; Brown, J.; Haggard, D. E.; Wambaugh, J.; Judson, R. S.; Paul Friedman, K. Estimating Uncertainty in the Context of New Approach Methodologies for Potential Use in Chemical Safety Evaluation. *Curr. Opin. Toxicol.* **2019**, *15*, 40–47. https://doi.org/https://doi.org/10.1016/j.cotox.2019.04.001.

(4) Isaacs, K. K.; Egeghy, P.; Dionisio, K. L.; Phillips, K. A.; Zidek, A.; Ring, C.; Sobus, J. R.; Ulrich, E. M.; Wetmore, B. A.; Williams, A. J.; Wambaugh, J. F. The Chemical Landscape of High-Throughput New Approach Methodologies for Exposure. *J. Expo. Sci. Environ. Epidemiol.* **2022**. https://doi.org/10.1038/s41370-022-00496-9.

(5) Pradeep, P.; Paul Friedman, K.; Judson, R. Structure-Based QSAR Models to Predict Repeat Dose Toxicity Points of Departure. *Comput. Toxicol.* **2020**, *16*, 100139. https://doi.org/https://doi.org/10.1016/j.comtox.2020.100139.

(6) Pham, L. L.; Watford, S. M.; Pradeep, P.; Martin, M. T.; Thomas, R. S.; Judson, R. S.; Setzer, R. W.; Paul Friedman, K. Variability in in Vivo Studies: Defining the Upper Limit of Performance for Predictions of Systemic Effect Levels. *Comput. Toxicol.* **2020**, *15*, 100126. https://doi.org/https://doi.org/10.1016/j.comtox.2020.100126.

(7) Blessinger, T.; Davis, A.; Chiu, W. A.; Stanek, J.; Woodall, G. M.; Gift, J.; Thayer, K.

A.; Bussard, D. Application of a Unified Probabilistic Framework to the Dose-Response Assessment of Acrolein. *Environ. Int.* **2020**, *143*, 105953. https://doi.org/https://doi.org/10.1016/j.envint.2020.105953.

(8) Jolliet, O.; Huang, L.; Hou, P.; Fantke, P. High Throughput Risk and Impact Screening of Chemicals in Consumer Products. *Risk Anal.* **2021**, *41* (4), 627–644. https://doi.org/https://doi.org/10.1111/risa.13604.

(9) Judson, R.; Ann, R.; J., D. D.; Keith, H.; Matthew, M.; Robert, K.; Vicki, D.; Tala, H.; Todd, H.; Philip, S.; Shirlee, T.; Thomas, C.; Edwin, S. The Toxicity Data Landscape for Environmental Chemicals. *Environ. Health Perspect.* **2009**, *117* (5), 685–695. https://doi.org/10.1289/ehp.0800168.

(10) Wang, Z.; Walker, G. W.; Muir, D. C. G.; Nagatani-Yoshida, K. Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. *Environ. Sci. Technol.* **2020**, *54* (5), 2575–2584. https://doi.org/10.1021/acs.est.9b06379.

(11) von Borries, K.; Holmquist, H.; Kosnik, M.; Beckwith, K. V; Jolliet, O.; Goodman, J. M.; Fantke, P. Potential for Machine Learning to Address Data Gaps in Human Toxicity and Ecotoxicity Characterization. *Environ. Sci. Technol.* **2023**, *57* (46), 18259–18270. https://doi.org/10.1021/acs.est.3c05300.

(12) Wignall, J. A.; Muratov, E.; Sedykh, A.; Guyton, K. Z.; Tropsha, A.; Rusyn, I.; Chiu, W. A. Conditional Toxicity Value (CTV) Predictor: An In Silico Approach for Generating Quantitative Risk Estimates for Chemicals. *Environ. Health Perspect.* **2018**, *126* (5), 057008. https://doi.org/10.1289/EHP2998.

(13) Chiu, W. A.; Paoli, G. M. Recent Advances in Probabilistic Dose–Response Assessment to Inform Risk-Based Decision Making. *Risk Anal.* **2021**, *41* (4), 596–609. https://doi.org/https://doi.org/10.1111/risa.13595.

(14) WHO. *Guidance Document on Evaluating and Expressing Uncertainty in Hazard*

Characterization.; Geneva, Switzerland, 2014.

(15) Chiu, W. A.; Slob, W. A Unified Probabilistic Framework for Dose–Response Assessment of Human Health Effects. *Environ. Health Perspect.* **2015**, *123* (12), 1241–1254. https://doi.org/10.1289/ehp.1409385.

(16) Chiu, W. A.; Axelrad, D. A.; Dalaijamts, C.; Dockins, C.; Shao, K.; Shapiro, A. J.; Paoli, G. Beyond the RfD: Broad Application of a Probabilistic Approach to Improve Chemical Dose–Response Assessments for Noncancer Effects. *Environ. Health Perspect.* **2018**, *126* (6), 067009. https://doi.org/10.1289/EHP3368.

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

(18) Aurisano, N.; Huang, L.; Milà i Canals, L.; Jolliet, O.; Fantke, P. Chemicals of Concern in Plastic Toys. *Environ. Int.* **2021**, *146*, 106194. https://doi.org/10.1016/j.envint.2020.106194.

(19) Aurisano, N.; Fantke, P.; Huang, L.; Jolliet, O. Estimating Mouthing Exposure to Chemicals in Children's Products. *J. Expo. Sci. Environ. Epidemiol.* **2022**, *32* (1), 94–102. https://doi.org/10.1038/s41370-021-00354-0.

(20) Huang, L.; Fantke, P.; Ritscher, A.; Jolliet, O. Chemicals of Concern in Building Materials: A High-Throughput Screening. *J. Hazard. Mater.* **2022**, *424*, 127574. https://doi.org/https://doi.org/10.1016/j.jhazmat.2021.127574.

(21) Huang, L.; Aurisano, N.; Fantke, P.; Dissanayake, A.; Edirisinghe, L. G. L. M.; Jolliet, O. Near-Field Exposures and Human Health Impacts for Organic Chemicals in Interior Paints: A High-Throughput Screening. *J. Hazard. Mater.* **2024**, *465*, 133145. https://doi.org/https://doi.org/10.1016/j.jhazmat.2023.133145.

(22) Judson, R. 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. 2019. https://doi.org/10.23645/epacomptox.7800653.v1.

(23) Nelms, M. D.; Patlewicz, G. Derivation of New Threshold of Toxicological Concern Values for Exposure via Inhalation for Environmentally-Relevant Chemicals . *Frontiers in Toxicology* . 2020. https://www.frontiersin.org/articles/10.3389/ftox.2020.580347.

(24) Huijbregts, M. A. J.; Rombouts, L. J. A.; Ragas, A. M. J.; van de Meent, D. Human-Toxicological Effect and Damage Factors of Carcinogenic and Noncarcinogenic Chemicals for Life Cycle Impact Assessment. *Integr. Environ. Assess. Manag.* **2005**, *1* (3), 181–244. https://doi.org/https://doi.org/10.1897/2004-007R.1.

(25) Jolliet, O.; Ernstoff, A. S.; Csiszar, S. A.; Fantke, P. Defining Product Intake Fraction to Quantify and Compare Exposure to Consumer Products. *Environ. Sci. Technol.* **2015**, *49* (15), 8924–8931. https://doi.org/10.1021/acs.est.5b01083.

(26) Fantke, P.; Ernstoff, A. S.; Huang, L.; Csiszar, S. A.; Jolliet, O. Coupled Near-Field and Far-Field Exposure Assessment Framework for Chemicals in Consumer Products. *Environ. Int.* **2016**, *94*, 508–518. https://doi.org/10.1016/j.envint.2016.06.010.

(27) Paul Friedman, K.; Gagne, M.; Loo, L.-H.; Karamertzanis, P.; Netzeva, T.; Sobanski, T.; Franzosa, J. A.; Richard, A. M.; Lougee, R. R.; Gissi, A.; Lee, J.-Y. J.; Angrish, M.; Dorne, J. Lou; Foster, S.; Raffaele, K.; Bahadori, T.; Gwinn, M. R.; Lambert, J.; Whelan, M.; Rasenberg, M.; Barton-Maclaren, T.; Thomas, R. S. Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization. *Toxicol. Sci.* **2020**, *173* (1), 202–225. https://doi.org/10.1093/toxsci/kfz201.

(28) Fantke, P.; Aurisano, N.; Provoost, J.; Karamertzanis, P. G.; Hauschild, M. Toward Effective Use of REACH Data for Science and Policy. *Environ. Int.* **2020**, *135* (November 2019), 105336. https://doi.org/10.1016/j.envint.2019.105336.

(29) Zeise, L.; Bois, F. Y.; Chiu, W. A.; Hattis, D.; Rusyn, I.; Guyton, K. Z. Addressing Human Variability in Next-Generation Human Health Risk Assessments of Environmental Chemicals. *Environ. Health Perspect.* **2013**, *121* (1), 23–31. https://doi.org/10.1289/ehp.1205687.

(30) Smith, M. N.; Cohen Hubal, E. A.; Faustman, E. M. 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.* **2019**, *30* (1), 160–170. https://doi.org/10.1038/s41370-019-0165-y.

(31) Aurisano, N.; Weber, R.; Fantke, P. Enabling a Circular Economy for Chemicals in Plastics. *Curr. Opin. Green Sustain. Chem.* **2021**, *31*, 100513. https://doi.org/https://doi.org/10.1016/j.cogsc.2021.100513.

(32) Aurisano, N.; Fantke, P. Semi-Automated Harmonization and Selection of Chemical Data for Risk and Impact Assessment. *Chemosphere* **2022**, *302*, 134886. https://doi.org/https://doi.org/10.1016/j.chemosphere.2022.134886.

(33) Aurisano, N.; Albizzati, P. F.; Hauschild, M.; Fantke, P. Extrapolation Factors for Characterizing Freshwater Ecotoxicity Effects. *Environ. Toxicol. Chem.* **2019**, *38* (11), 2568–2582. https://doi.org/10.1002/etc.4564.

(34) Williams, A. J.; Grulke, C. M.; Edwards, J.; McEachran, A. D.; Mansouri, K.; Baker, N. C.; Patlewicz, G.; Shah, I.; Wambaugh, J. F.; Judson, R. S.; Richard, A. M. The CompTox Chemistry Dashboard: A Community Data Resource for Environmental Chemistry. *J. Cheminform.* **2017**, *9* (1), 1–27. https://doi.org/10.1186/s13321-017-0247-6.

(35) Williams, A. J.; Lambert, J. C.; Thayer, K.; Dorne, J.-L. C. M. Sourcing Data on

749 Chemical Properties and Hazard Data from the US-EPA CompTox Chemicals
750 Dashboard: A Practical Guide for Human Risk Assessment. *Environ. Int.* **2021**, *154*,
751 106566. https://doi.org/https://doi.org/10.1016/j.envint.2021.106566.
752 (36) Martin, M. T.; Judson, R. S.; Reif, D. M.; Kavlock, R. J.; Dix, D. J. Profiling
753 Chemicals Based on Chronic Toxicity Results from the U.S. EPA ToxRef Database.
754 *Environ. Health Perspect.* **2009**, *117* (3), 392–399.
755 https://doi.org/10.1289/ehp.0800074.
756 (37) Watford, S.; Ly Pham, L.; Wignall, J.; Shin, R.; Martin, M. T.; Friedman, K. P.
757 ToxRefDB Version 2.0: Improved Utility for Predictive and Retrospective Toxicology
758 Analyses. *Reprod. Toxicol.* **2019**, *89*, 145–158.
759 https://doi.org/10.1016/j.reprotox.2019.07.012.
760 (38) Kosnik, M. B.; Kephalopoulos, S.; Muñoz, A.; Aurisano, N.; Cusinato, A.;
761 Dimitroulopoulou, S.; Slobodnik, J.; De Mello, J.; Zare Jeddi, M.; Cascio, C.; Ahrens,
762 A.; Bruinen de Bruin, Y.; Lieck, L.; Fantke, P. Advancing Exposure Data Analytics
763 and Repositories as Part of the European Exposure Science Strategy 2020–2030.
764 *Environ. Int.* **2022**, *170*, 107610.
765 https://doi.org/https://doi.org/10.1016/j.envint.2022.107610.
766 (39) Li, L.; Zhang, Z.; Men, Y.; Baskaran, S.; Sangion, A.; Wang, S.; Arnot, J. A.; Wania,
767 F. Retrieval, Selection, and Evaluation of Chemical Property Data for Assessments of
768 Chemical Emissions, Fate, Hazard, Exposure, and Risks. *ACS Environ. Au* **2022**, *2* (5),
769 376–395. https://doi.org/10.1021/acsenvironau.2c00010.
770 (40) Wignall, J. A.; Shapiro, J. A.; Wright, A. F.; Woodruff, J. T.; Chiu, A. W.; Guyton, Z.
771 K.; Rusyn, I. Standardizing Benchmark Dose Calculations to Improve Science-Based
772 Decisions in Human Health Assessments. *Environ. Health Perspect.* **2014**, *122* (5),
773 499–505. https://doi.org/10.1289/ehp.1307539.
774 (41) US Environmental protection Agency. *Regional Screening Levels (RSLs) - Generic*

Tables. https://www.epa.gov/risk/regional-screening-levels-rsls-generic-tables (accessed 2020-03-17).

(42) Posthuma, L.; van Gils, J.; Zijp, M. C.; van de Meent, D.; de Zwartd, D. Species Sensitivity Distributions for Use in Environmental Protection, Assessment, and Management of Aquatic Ecosystems for 12 386 Chemicals. *Environ. Toxicol. Chem.* **2019**, *38* (4), 905–917. https://doi.org/10.1002/etc.4373.

(43) Fantke, P.; Wieland, P.; Juraske, R.; Shaddick, G.; Itoiz, E. S.; Friedrich, R.; Jolliet, O. Parameterization Models for Pesticide Exposure via Crop Consumption. *Environ. Sci. Technol.* **2012**, *46* (23), 12864–12872. https://doi.org/10.1021/es301509u.

(44) Hong, J.; Shaked, S.; Rosenbaum, R. K.; Jolliet, O. Analytical Uncertainty Propagation in Life Cycle Inventory and Impact Assessment: Application to an Automobile Front Panel. *Int. J. Life Cycle Assess.* **2010**, *15* (5), 499–510. https://doi.org/10.1007/s11367-010-0175-4.

(45) Stylianou, K. S.; Fulgoni, V. L.; Jolliet, O. Small Targeted Dietary Changes Can Yield Substantial Gains for Human and Environmental Health. *Nat. Food* **2021**, *2* (8), 616–627. https://doi.org/10.1038/s43016-021-00343-4.

(46) Slob, W. Uncertainty Analysis in Multiplicative Models. *Risk Anal.* **1994**, *14* (4), 571–576. https://doi.org/https://doi.org/10.1111/j.1539-6924.1994.tb00271.x.

(47) Ring, C. L.; Arnot, J. A.; Bennett, D. H.; Egeghy, P. P.; Fantke, P.; Huang, L.; Isaacs, K. K.; Jolliet, O.; Phillips, K. A.; Price, P. S.; Shin, H. M.; Westgate, J. N.; Setzer, R. W.; Wambaugh, J. F. Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways. *Environ. Sci. Technol.* **2019**, *53* (2), 719–732. https://doi.org/10.1021/acs.est.8b04056.

(48) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. https://www.r-project.org/.

(49) Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag New York, 2016.

(50) Browne, P.; Judson, R. S.; Casey, W. M.; Kleinstreuer, N. C.; Thomas, R. S. Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model. *Environ. Sci. Technol.* **2015**, *49* (14), 8804–8814. https://doi.org/10.1021/acs.est.5b02641.

(51) Kleinstreuer, C. N.; Ceger, C. P.; Allen, G. D.; Strickland, J.; Chang, X.; Hamm, T. J.; Casey, M. W. A Curated Database of Rodent Uterotrophic Bioactivity. *Environ. Health Perspect.* **2016**, *124* (5), 556–562. https://doi.org/10.1289/ehp.1510183.

(52) Gramatica, P. On the Development and Validation of QSAR Models BT - Computational Toxicology: Volume II; Reisfeld, B., Mayeno, A. N., Eds.; Humana Press: Totowa, NJ, 2013; pp 499–526. https://doi.org/10.1007/978-1-62703-059-5_21.

(53) Office of Environmental Health Hazard Assessment (OEHHA). Technical Support Document for the Derivation of Noncancer Reference Exposure Levels. Air Toxic Hot Spots, Risk Assessment Guidelines. 2008. http://www.oehha.ca.gov/air/hot_spots/rels_dec2008.html.

(54) Rosenbaum, R. K.; Bachmann, T. M.; Gold, L. S.; Huijbregts, M. A. J.; Jolliet, O.; Juraske, R.; Koehler, A.; Larsen, H. F.; MacLeod, M.; Margni, M.; McKone, T. E.; Payet, J.; Schuhmacher, M.; Van De Meent, D.; Hauschild, M. Z. USEtox - The UNEP-SETAC Toxicity Model: Recommended Characterisation Factors for Human Toxicity and Freshwater Ecotoxicity in Life Cycle Impact Assessment. *Int. J. Life Cycle Assess.* **2008**, *13* (7), 532–546. https://doi.org/10.1007/s11367-008-0038-4.

(55) Hauschild, M. Z.; Huijbregts, M.; Jolliet, O.; Macleod, M.; Margni, M.; Van De Meent, D.; Rosenbaum, R. K.; McKone, T. E. Building a Model Based on Scientific Consensus for Life Cycle Impact Assessment of Chemicals: The Search for Harmony and Parsimony. *Environmental Science and Technology*. 2008, pp 7032–7037. https://doi.org/10.1021/es703145t.

(56) Owsianiak, M.; Hauschild, M. Z.; Posthuma, L.; Saouter, E.; Vijver, M. G.; Backhaus, T.; Douziech, M.; Schlekat, T.; Fantke, P. Ecotoxicity Characterization of Chemicals: Global Recommendations and Implementation in USEtox. *Chemosphere* **2023**, *310*, 136807. https://doi.org/https://doi.org/10.1016/j.chemosphere.2022.136807.

(57) Rosenbaum, R. K.; Bachmann, T. M.; Gold, L. S.; Huijbregts, M. A. J.; Jolliet, O.; Juraske, R.; Koehler, A.; Larsen, H. F.; MacLeod, M.; Margni, M.; McKone, T. E.; Payet, J.; Schuhmacher, M.; Van De Meent, D.; Hauschild, M. Z. USEtox - The UNEP-SETAC Toxicity Model: Recommended Characterisation Factors for Human Toxicity and Freshwater Ecotoxicity in Life Cycle Impact Assessment. *Int. J. Life Cycle Assess.* **2008**, *13* (7), 532–546. https://doi.org/10.1007/s11367-008-0038-4.

(58) Rosenbaum, R. K.; Huijbregts, M. A. J. J.; Henderson, A. D.; Margni, M.; McKone, T. E.; Van De Meent, D.; Hauschild, M. Z.; Shaked, S.; Li, D. S.; Gold, L. S.; Jolliet, O. USEtox Human Exposure and Toxicity Factors for Comparative Assessment of Toxic Emissions in Life Cycle Analysis: Sensitivity to Key Chemical Properties. *Int. J. Life Cycle Assess.* **2011**, *16* (8), 710. https://doi.org/10.1007/s11367-011-0316-4.

(59) National Research Council (NRC). *Science and Decisions: Advancing Risk Assessment*; The National Academies Press: Washington, DC, 2009. https://doi.org/10.17226/12209.

(60) Emara, Y.; Fantke, P.; Judson, R.; Chang, X.; Pradeep, P.; Lehmann, A.; Siegert, M.-W.; Finkbeiner, M. Integrating Endocrine-Related Health Effects into Comparative Human Toxicity Characterization. *Sci. Total Environ.* **2021**, *762*, 143874. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.143874.

(61) Mansouri, K.; Karmaus, A. L.; Fitzpatrick, J.; Patlewicz, G.; Pradeep, P.; Alberga, D.; Alepee, N.; Allen, T. E. H.; Allen, D.; Alves, V. M.; Andrade, C. H.; Auernhammer, T. R.; Ballabio, D.; Bell, S.; Benfenati, E.; Bhattacharya, S.; Bastos, J. V.; Boyd, S.; Brown, J. B.; Capuzzi, S. J.; Chushak, Y.; Ciallella, H.; Clark, A. M.; Consonni, V.;

Daga, P. R.; Ekins, S.; Farag, S.; Fedorov, M.; Fourches, D.; Gadaleta, D.; Gao, F.; Gearhart, J. M.; Goh, G.; Goodman, J. M.; Grisoni, F.; Grulke, C. M.; Hartung, T.; Hirn, M.; Karpov, P.; Korotcov, A.; Lavado, G. J.; Lawless, M.; Li, X.; Luechtefeld, T.; Lunghini, F.; Mangiatordi, G. F.; Marcou, G.; Marsh, D.; Martin, T.; Mauri, A.; Muratov, E. N.; Myatt, G. J.; Nguyen, D. T.; Nicolotti, O.; Note, R.; Pande, P.; Parks, A. K.; Peryea, T.; Polash, A. H.; Rallo, R.; Roncaglioni, A.; Rowlands, C.; Ruiz, P.; Russo, D. P.; Sayed, A.; Sayre, R.; Sheils, T.; Siegel, C.; Silva, A. C.; Simeonov, A.; Sosnin, S.; Southall, N.; Strickland, J.; Tang, Y.; Teppen, B.; Tetko, I. V.; Thomas, D.; Tkachenko, V.; Todeschini, R.; Toma, C.; Tripodi, I.; Trisciuzzi, D.; Tropsha, A.; Varnek, A.; Vukovic, K.; Wang, Z.; Wang, L.; Waters, K. M.; Wedlake, A. J.; Wijeyesakere, S. J.; Wilson, D.; Xiao, Z.; Yang, H.; Zahoranszky-Kohalmi, G.; Zakharov, A. V.; Zhang, F. F.; Zhang, Z.; Zhao, T.; Zhu, H.; Zorn, K. M.; Casey, W.; Kleinstreuer, N. C. CATMoS: Collaborative Acute Toxicity Modeling Suite. *Environ. Health Perspect.* **2021**, *129* (4), 1–18. https://doi.org/10.1289/EHP8495.

(62) Hou, P.; Jolliet, O.; Zhu, J.; Xu, M. Estimate Ecotoxicity Characterization Factors for Chemicals in Life Cycle Assessment Using Machine Learning Models. *Environ. Int.* **2020**, *135*, 105393. https://doi.org/10.1016/j.envint.2019.105393.