Towards a quantitative adverse outcome pathway for liver carcinogenesis: from proliferation to prediction

Veltman, C.H.J.; Khalidi, H.; Zgheib, E.; Water, B. van de; Luijten, M.; Pennings, J.L.A.

Computational Toxicology 34 (2025) 100359

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.sciencedirect.com/journal/computational-toxicology

Full Length Article

Towards a quantitative adverse outcome pathway for liver carcinogenesis: From proliferation to prediction

Christina H.J. Veltman [a,b,*], Hiba Khalidi [c], Elias Zgheib [c], Bob van de Water [b], Mirjam Luijten [a], Jeroen L.A. Pennings [a]

[a] Centre for Health Protection, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
[b] Division of Drug Discovery and Safety, Leiden Academic Center for Drug Research (LACDR), Leiden University, Leiden, the Netherlands
[c] Division Simcyp, Certara Predictive Technologies, Sheffield, United Kingdom

ARTICLE INFO

Keywords:
Carcinogenesis
Response-Response modelling
Bayesian logistic regression
Quantitative Adverse Outcome Pathway (qAOP)
Key Event Relationship (KER)

ABSTRACT

Hazard assessment of non-genotoxic carcinogens could greatly benefit from next generation risk assessment approaches, driven by the multitude of mechanisms through which non-genotoxic carcinogens operate. One method for structuring new approach methodology-derived data is the adverse outcome pathway (AOP) concept. Currently, mostly qualitative AOPs are described, limiting their application for regulatory decision making. In contrast, quantitative AOPs use mathematical terms to describe the relationships between key events (KEs), allowing for the derivation of a Point of Departure (PoD). Here, we report quantification of the key event relationship (KER) between sustained hepatocyte proliferation and liver tumour formation, two KEs of AOP#220 relating to CYP2E1 activation leading to liver cancer. We use incidence of histopathological lesions indicative of proliferation, as well as BrdU labelling obtained from existing sub-chronic toxicity studies in rats, to quantify proliferation. For liver cancer, incidences of hepatocellular adenoma and carcinoma from 2-year rodent carcinogenicity studies were collected. Data for both KEs were combined to calibrate a response-response model, and Bayesian logistic regression analysis was applied to obtain predictions and credible intervals for carcinogenicity. Proliferative lesion incidence was observed to be a highly specific, yet insensitive predictor, and combining this with BrdU labelling yields more accurate predictions of carcinogenicity. Importantly, we demonstrate that for most of the chemicals tested, inclusion of BrdU labelling returns more precise predicted benchmark dose intervals for PoD derivation. To further explore this quantitative KER and its regulatory application, we propose to include and standardize BrdU labelling for sub-chronic toxicity studies performed for regulatory purposes.

1. Introduction

Carcinogenesis is a complex multi-step process in which normal cells are transformed into cancer cells [1]; as such it is considered an essential element in the human health risk assessment of chemicals [2,3]. Carcinogenic chemicals can be subdivided into two main categories: genotoxic or direct DNA interacting carcinogens, and non-genotoxic or non-direct DNA interacting carcinogens [4,5]. Traditional hazard assessment of carcinogenic chemicals involves the two-year rodent cancer bioassays, as described in OECD Test Guidelines 451 and 453 [6,7]. However, these studies have equivocal reproducibility [8] and raise ethical concerns [9]. As genetic damage is considered key to carcinogenesis, novel strategies often evolve around genotoxic endpoints under both *in vivo* and *in vitro* conditions [10]. However, these tests are not designed to detect non-genotoxic carcinogens [11]. As a result, there is an increasing demand for novel approaches in cancer hazard assessment, based on mechanistic insights rather than on methods based on apical endpoints.

Within toxicology, adverse outcome pathways (AOPs) are used to structure the biological key events leading to an adverse outcome in a chemical-independent manner [12,13]. Additionally, AOPs support the integration of new approach methodologies (NAMs) for chemical hazard assessment [14,15]. Current AOP development is mostly focused on qualitative AOPs, limiting their application in regulatory decision making, as quantitative differences are often key in discerning an adaptive from an adverse response in an exposed organism [16,17]. To address this, quantitative AOPs (qAOPs) describe the relationship between key events in mathematical terms [18,19,17]. This allows for extrapolation of specific (early) key event perturbations to an adverse

* Corresponding author.
E-mail address: kirsten.veltman@rivm.nl (C.H.J. Veltman).

https://doi.org/10.1016/j.comtox.2025.100359
Received 28 March 2025; Received in revised form 29 May 2025; Accepted 29 May 2025
Available online 1 June 2025

outcome probability or severity prediction [19]. As a result, qAOPs can be used for benchmark dose (BMD) modelling and selecting a point of departure (PoD) for the derivation of safe levels. Methods for AOP quantification include effect size modelling (response-response; [20,21,22,23,24]), probabilistic modelling (frequentists and Bayesian regression; [25,26,27,28,29]), and mechanistic modelling (i.e. systems toxicology; [30]). Despite increasing interest, there are currently no guidelines on how to make best use of qAOPs in hazard assessment [12,31].

In the past decade, a substantial number of AOPs for carcinogenesis have been proposed [32,33]. Regardless of the multitude of mechanisms involved in non-genotoxic carcinogenicity and their diversity in target organs, increased cell proliferation is considered a fundamental (late) key event in carcinogenesis [34,35]. This is reflected in the AOPwiki, a repository for proposed AOPs, which contains 40 AOPs describing both increased cell proliferation and tumour formation [32]. Nonetheless, these AOPs lack the quantitative aspect needed for uptake in regulatory decision making. As the key event "increased cell proliferation" is shared amongst various modes of action for non-genotoxic carcinogens, we consider this key event to be a relevant starting point for quantitative AOP development. Since there are currently no OECD validated *in vitro* cell proliferation assays to report on carcinogenesis [35], we reverted to sub-chronic repeated dose toxicity studies to obtain relevant information.

Predicting (absence of) carcinogenicity based on sub-chronic repeated dose toxicity studies is not new. Multiple studies have focused on sub-chronic indicators of carcinogenicity, including hyperplasia, hypertrophy, increased organ weight and tissue degeneration, and showed that these indicators were not sufficient to reliably predict carcinogenicity [36,37,38,39]. Another approach, referred to as Negative for Endocrine, Genotoxicity, and Chronic Study Associated Histopathologic Risk Factors for Carcinogenicity (NegCarc), focused on predicting absence of carcinogenicity based on absence of preneoplastic lesions in sub-chronic repeated dose toxicity studies, in combination with absence of genotoxic and hormonal disturbance potential [40,41,42,43]. Specific proliferation detection methods, such as bromodeoxyuridine (BrdU), Proliferating Cell Nuclear Antigen (PCNA) and Ki-67 labelling, were suggested to allow for more enhanced detection of proliferative effects compared to histopathological findings [44], with BrdU suggested to be the most robust marker of the three [45]. BrdU labelling was shown to have reasonable accuracy in predicting (absence of) carcinogenicity after short-term exposure [46,47], with specificity increasing and sensitivity decreasing with time [38]. Nevertheless, the products of these approaches are qualitative and can therefore not be used to derive a PoD.

In this study, we report quantification of the key event relationship (KER) between sustained hepatocyte proliferation and liver tumour formation, two (late) key events of, amongst others, the Organisation for Economic Cooperation and Development (OECD) endorsed AOP on CYP2E1 activation leading to liver cancer (Fig. 1; [48]). Our objective was to predict the probability of liver tumour induction and obtain PoDs for risk assessment purposes, based on existing *in vivo* proliferation data obtained from sub-chronic repeated dose toxicity studies in rats. We used a subset of five histopathological lesions (hepatocyte hyperplasia, multinucleated hepatocytes, basophilic foci, acidophilic foci and mitotic alterations), as well as BrdU labelling, to quantify dose–response relationships for sustained proliferation. For the adverse outcome, incidences of hepatocellular tumours (adenoma and carcinoma) from 2-year rodent carcinogenicity studies were collected for dose–response modelling. We used publicly available data for both key events to calibrate a response-response model and applied multivariable Bayesian logistic regression analysis to obtain a predicted BMD and credible interval for liver tumorigenicity based on cellular proliferation. We report on the results from the different modelling approaches and discuss strengths, weaknesses and further research needs.

2. Materials and methods

To facilitate understanding of the methodological approach taken in this study, the workflow is illustrated in Fig. 2. This schematic provides an overview of the key steps, which are described in more detail in the corresponding sections. The study began with data collection from the public domain (2.1 Data collection and processing). After data collection, dose–response curves were fitted for data imputation (2.2 Dose-response modelling). Next, response-response modelling (2.3 Response-response modelling) and Bayesian logistic regression (2.4 Bayesian logistic regression) were performed in parallel. To account for toxicokinetic differences between chemicals, response-response modelling was also performed using data corrected for internal liver concentrations (2.6 Internal dose level estimation). Multivariable Bayesian logistic regression (2.4 Bayesian logistic regression) was then conducted. Finally, both single and multivariable Bayesian logistic regression models were used for benchmark dose (BMD) analysis, with the resulting BMDs compared to those derived from 2-year cancer bioassays (2.5 Benchmark dose analysis).

2.1. Data collection and processing

Data from sub-chronic toxicity and 2-year carcinogenicity studies were extracted from National Toxicology Program (NTP) technical reports, Joint Meeting on Pesticide Residues (JMPR) regulatory assessment reports, European Food Safety Authority (EFSA) draft assessment reports, and scientific publications (searched between February 2023 and March 2024; supplementary references). BrdU labelling and proliferative histopathological lesions (hepatocyte hyperplasia, multinucleated hepatocytes, basophilic foci, acidophilic foci, and mitotic alterations) were selected as markers for sustained proliferation, the latter based on expert opinion. The following criteria were set for data selection: (1) data generated in rats, (2) route of exposure either oral (diet, gavage and drinking water) or inhalation, (3) approximately 90 days (range: 84 – 98) of chemical exposure, (4) at least three dose levels in addition to the vehicle control, and (5) BrdU administration of at least three days (range: five to seven) at the end of the study (BrdU only) and presence of any histopathological lesion in the liver at 90 days (histopathology only; used to reduce the imbalance toward negative data entries). For the resulting chemicals, 2-year carcinogenicity data reporting hepatocellular adenoma and carcinoma were collected. Criteria for carcinogenicity data include: (1) at least three dose levels, (2) commonality of strain and sex compared to BrdU or histopathology data, and (3) overlap in the tested dose-range between the sub-chronic and chronic study. Chemicals were considered liver carcinogens if they exhibited a dose-dependent increase in liver tumour incidence, with at least one dose group showing a statistically significant effect ($p < 0.05$). Data for liver non-carcinogens were taken along as well to

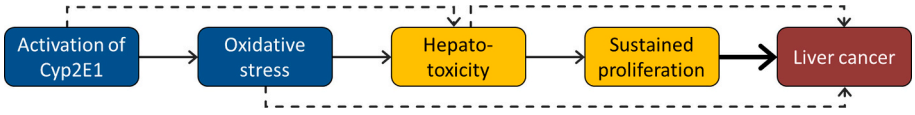


Fig. 1. Linear AOP for CYP2E1 induction leading to liver tumour formation (AOP#220). Solid lines represent direct key event relationships (KERs) and dashed lines represent indirect KERs. Blue boxes represent molecular and cellular key events, yellow boxes represent tissue-level key events, and red represents the adverse outcome. Adapted from [48].

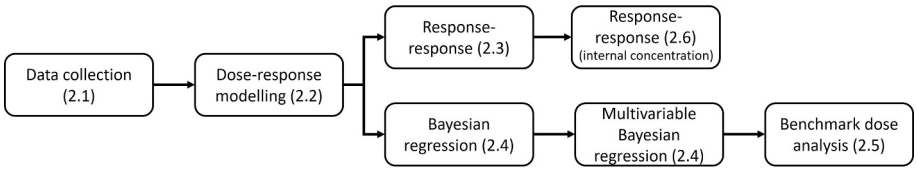


Fig. 2. Workflow of the methodological approach. Schematic overview of the key steps of the study presented in sequential order as labelled boxes. Detailed descriptions of each step are provided in the corresponding sections of the Materials and methods, as indicated in parentheses.

assess model performance. Additionally, studies were assessed for compliance with OECD test guidelines 408, 413, 451 or 453 [6,7,49,50]. Exemptions on the number of animals per dose group or number of chemical dose levels were allowed for three chemicals. The resulting overview of chemicals is listed in supplementary Table S1.

Data from various study reports were manually extracted and collected into an Excel spreadsheet. Histopathological proliferative lesion incidence data were pre-processed to obtain the combined overall incidence of all five lesions of interest, counting animals with two or more different lesions only once. Severity of the lesions was not taken into account. BrdU labelling indices were transformed to estimates of the total number of BrdU labelled cells by multiplying the reported percentage by the total number of counted cells.

2.2. Dose-response modelling

Dose-response curves for the collected chemicals were generated using PROAST (version 70.3 [51,52]; supplementary Table S2). Tumour incidence, proliferative lesion incidence and BrdU labelling data were considered quantal, therefore a log-logistic function (function 1) was chosen. Model parameters of fitted dose–response curves are listed in supplementary Table S3. To maximize overlap between the different datasets, the generated dose–response curves for tumour incidence were used for data imputation for doses administered in the corresponding sub-chronic toxicity study. For carbon tetrachloride, dose–response curves for BrdU labelling and proliferative lesion incidence were used for data imputation in the combined dataset, as tested doses in the respective sub-chronic toxicity studies did not overlap.

Function 1 – Log-logistic response curve:

$y = a + \frac{1-a}{1+\exp(c * \ln(\frac{b}{x}))}$ where x is the administered dose and y is the response. Variables a, b, and c are fit-dependent constants representing the background response, the inflection point (halfway response), and the slope, respectively.

2.3. Response-response modelling

For response-response modelling, dose matched conditions in both sub-chronic and chronic toxicity studies for each individual liver carcinogen were fitted in PROAST using a log-logistic function (function 1). Uncertainty in the fit was captured by bootstrapping 500 times and calculating the 95 % confidence interval of the fitted response-response curve. To model a chemical-agnostic response-response relationship, data were first normalized to the corresponding vehicle control. For BrdU, data were normalized using the fold change (function 2). Extra risk (function 3), defined as the scaled proportion of the total risk corrected for background risk, was used to normalized both tumour and proliferative lesion incidence data [53]. Secondly, a chemical-agnostic response-response curve was generated by taking the average of the normalized chemical-specific relations, and the 95 % confidence interval of the fit was calculated. For comparison, a chemical-agnostic response-response curve was also directly fitted in PROAST using the normalized values. Data used for response-response modelling can be found in supplementary Tables S4-S6.

Carcinogenicity thresholds for proliferative lesion incidence and fold change BrdU labelling were determined by taking a critical effect size of 10 % extra tumour incidence [53], and calculating the intersection between this and the chemical-agnostic response-response curve.

Function 2 – Fold change (FC):

$FC = \frac{x}{a}$ where x is the modelled BrdU labelling and a is the background BrdU labelling.

Function 3 – Extra risk (ER):

$ER = \frac{x-a}{1-a}$ where x is the modelled incidence and a is the background incidence.

2.4. Bayesian logistic regression

A Bayesian logistic regression model was developed using the brm function (brms package [54], version 2.19.0) in R (version 4.3.2). As before, we applied a threshold of 10 % extra tumour incidence as the critical effect size for carcinogenicity. Predictivity of the proliferative lesions was compared as a continuous (incidence) and categorical (presence or absence) variable. For BrdU, a comparison between a continuous (fold change) and a categorical (above or below the response-response based threshold) variable model was made as well. BrdU data were log-transformed because of the right-skewed distribution of the data. Default priors were used for the single-variable models. Markov Chain Monte Carlo (MCMC) No-U-Turn sampler (NUTS) was used for posterior sampling, where 6,000 samples per chain were drawn from four independent chains with the first third being used for warmup. All models showed satisfactory convergence, depicted as convergence diagnostic values (Rhat) of 1.00. No divergent transitions were observed for any of the models, and effective sample sizes were $> 5{,}000$. Models were compared based on the Bayes factor (BF; bayes_factor function, brms package), area under the curve (AUC; auc function, pROC package [55], version 1.18.5), log marginal likelihood (bridge_sampler function, brms package), and Leave-One-Out cross validation on chemical level (loo function, loo package [56,57], version 2.7.0). The multivariable models were fitted on the subset of the data containing information on both variables. Their priors were informed by the posterior distributions from the corresponding single-variable models, which were based on the complete dataset available for each individual variable. Data used for Bayesian logistic regression modelling can be found in supplementary Tables S7-S9.

Diagnostic performance of the models was assessed based on accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Due to the absence of an independent dataset, model performance was evaluated on the same data using Leave-One-Out cross validation on the chemical level. In short, training datasets were made by taking the whole dataset except for one chemical, which was placed in the corresponding testing dataset. For each training dataset, a model was trained and predictions for the left-out chemical were made by taking the mean of 1,000 draws from the posterior distribution. All predictions for left-out chemicals were combined in one dataset, which was used for calculating model performance. A probability cut-off of 0.5 was used to assign probabilities into positives or negatives.

Function 4 – Accuracy:

$$\text{Accuracy} = \frac{\text{true positive} + \text{true negative}}{\text{total}}$$

Function 5 – Sensitivity:

$$\text{Sensitivity} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

Function 6 – Specificity:

$$\text{Specificity} = \frac{\text{true negative}}{\text{true negative} + \text{false positive}}$$

Function 7 – Positive predictive value (PPV):

$$\text{PPV} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

Function 8 – Negative predictive value (NPV):

$$\text{NPV} = \frac{\text{true negative}}{\text{true negative} + \text{false negative}}$$

2.5. Benchmark dose (BMD) analysis

BMD analysis was performed in R using PROAST. In short, probability predictions for 2-year liver carcinogenicity were obtained from the respective Bayesian regression models by taking the mean and credible interval of 1,000 posterior draws. Next, dose–response curves with predicted probabilities as response were fit using a log-logistic model (function 1) in PROAST. The carcinogenicity BMD was determined for a critical effect size of 50 %, representing a predicted probability of 50 % for carcinogenicity (an extra tumour incidence of 10 % or higher). BMD intervals were determined based on fitted log-logistic models for the credible intervals obtained by Bayesian regression modelling. For comparison, the reported 2-year tumour incidences were fit using a log-logistic function (function 1) in PROAST and BMD intervals were determined for a critical effect size of 10 % extra tumour incidence and based on 1,000 bootstraps. Finally, predicted probability based BMD intervals were compared to BMD intervals based on 2-year carcinogenicity data.

2.6. Internal dose level estimation

Internal liver concentrations following oral exposure were not reported for some chemicals. Therefore, for these chemicals, physiologically-based kinetic (PBK) modelling was performed using the rat model of the Simcyp animal simulator (version 23; [58]). The data required for PBK modelling were obtained from various sources, including published literature and predictions on physicochemical and pharmacokinetic properties.

- **Physicochemical properties:** Molecular weight, the logarithm of the octanol:water partition coefficient (LogP), polar surface area, number of hydrogen bond donors, chemical type, and pKa values were obtained from ChEMBL [59].
- **Blood Binding properties:** fraction unbound (f_u), and blood to plasma ratio (BP) parameters were mostly obtained from the literature. If no data were present in the literature, either predictions from the Simcyp Simulator or the ADMET predictor [60] were used instead.
- **Absorption:** The effective permeability (P_{eff}) of the chemicals was predicted by the Simcyp simulator using either (A) the Mechanistic P_{eff} model, or (B) the polar surface area and number of hydrogen bond donors. Either the advanced dissolution, absorption and metabolism (ADAM; [61]) model or first order absorption models were applied to predict the behaviour of absorption.
- **Distribution:** For all chemicals, a full PBK model was generated. The method of prediction of volume of distribution (V_{ss}) was selected based on chemical properties.
- **Elimination**: The elimination of each chemical was modelled based on the available literature data, incorporating multiple possible routes, including hepatic, renal, biliary, and enzymatic clearance (Table 1). The selection of elimination pathways for a given compound was determined by the available experimental and computational data, allowing for the possibility of multiple concurrent elimination mechanisms.
  - *In vivo* clearance data: Experimentally observed renal and biliary clearance in rats was directly implemented in the PBK models to replicate *in vivo* elimination.
  - *In silico* clearance predictions: Predicted hepatic intrinsic clearance values were used as the primary elimination pathway in the PBK models when *in vivo* data were unavailable.
  - Enzymatic clearance data: For compounds with reported Michaelis-Menten kinetic parameters (V_{max} and K_m) for relevant metabolic enzymes, enzymatic clearance was explicitly incorporated into the PBK models.

The models built were simulated at the desired dose levels (supplementary files 1–3), and the simulated blood/plasma concentration–time profiles were validated against measured concentration profiles in rat populations reported in literature. Simulated rat liver concentration profiles were validated against measured rat liver concentration profiles where available. Each chemical was validated on its respective set of measured data for plasma/liver profiles and/or clearance profiles from varying sources where available: hexachlorobenzene [63,66], methapyrilene [67,68], and isoxaflutole [65]. Table 1 summarises the main parameters used to build the PBK models for the chemicals of interest.

For carbon tetrachloride, an inhalation PBK model was developed in R (version 4.4.1) by integrating and refining previously published models [69–73]. The model was structured as a system of ordinary differential equations (ODEs) describing the absorption, distribution, metabolism, and elimination (ADME) of carbon tetrachloride across multiple physiological compartments. These compartments included (1) an exposure chamber, whose volume was adjusted based on the number of animals in each experiment to capture realistic gas exchange dynamics, (2) the lung, where inhaled carbon tetrachloride enters systemic circulation, (3) fat, taking into account any potential storage of the compound, (4) richly perfused tissues (e.g., kidney, viscera), (5) poorly perfused tissues, and (6) the liver, where metabolism follows Michaelis-

Table 1
Summary of parameters used to parametrize the rat model in v23 of the Simcyp animal simulator.

Chemical	Blood Binding	Absorption	Distribution	Elimination
Hexachlorobenzene LogP = 5.61 [59]	f_u: 0.027 [60] BP: 4.79 [62]	ADAM Model P_{eff} predicted using the Mechanistic P_{eff} model	Method 1 prediction for V_{ss}	Enzymatic Clearance (metabolising enzymes in the liver) Biliary and Renal Clearance (*in vivo*; [63])
Methapyrilene LogP = 3.11 [59]	f_u: 0.436 BP: 1.248 [60]	ADAM Model P_{eff} predicted using the Mechanistic P_{eff} model	Method 2 prediction for V_{ss}	Hepatic intrinsic clearance (*in silico*; [60])
Isoxaflutole LogP = 2.14 [59]	f_u: 0.062 [64] BP: 0.923 [58]	First order absorption Model P_{eff} predicted using the Mechanistic P_{eff} model	Method 1 prediction for V_{ss}	Hepatic intrinsic clearance (*in silico*; [60]) Biliary and Renal Clearance (*in vivo*, [65])

Menten kinetics. In this model, the term "liver" encompasses all tissues where metabolic clearance occurs. Simulated dose levels are listed in supplementary file 4. Simulated internal liver concentrations were used for response-response modelling.

3. Results

3.1. Data selection yields a limited but relatively harmonized dataset

The search for public data related to liver BrdU labelling and proliferative histopathological lesions identified 65 and 243 unique study reports, respectively (Fig. 3). Following screening according to the set data criteria, 20 reports were available for BrdU labelling and 69 for proliferative lesions. After coupling these data to carcinogenicity studies, 14 and 50 reports remained respectively. Together, these reports contain data on 51 chemicals, largely overlapping between the two datasets (Fig. 3). For BrdU labelling, data were collected for 10 liver carcinogens and 3 liver non-carcinogens. The distribution for proliferative lesions was 16 liver carcinogens and 34 liver non-carcinogens. Focussing on the overlapping chemicals for a combined dataset, we ended up with 10 liver carcinogens and 2 liver non-carcinogens. The list of resulting chemicals can be found in supplementary Table S1.

The collected data were subjected to various analyses (Table 2); each of these is described in more detail below. At first, we generated a response-response model for tumour incidence based on the incidence of proliferative lesions detected after sub-chronic exposure (Table 2; analysis #1) to obtain a relatively straightforward relationship focusing on the magnitude of effect. The resulting response-response curve was used to determine a threshold in proliferative lesion incidence for carcinogenicity predictions. Next, Bayesian modelling, which allows to deal with uncertainty and is considered more suitable for sparse datasets, was used in combination with the obtained threshold in a categorical logistic regression model (Table 2; analysis #2). For comparison, we also fitted a continuous Bayesian model based on proliferative lesion incidence (Table 2; analysis #3). Since proliferative lesions provide a semi-quantitative predictor, we also developed similar models for the quantitative predictor BrdU labelling. Again, we started with response-response modelling (Table 2; analysis #4) to gain insight into the relationship and determine a carcinogenicity threshold. Additionally, we checked if differences in administered doses and internal liver concentrations influenced the response-response relation (Table 2; analysis #5). Similar to proliferative lesions, we developed a categorical Bayesian logistic regression model based on the response-response obtained threshold (Table 2; analysis #6), and a continuous Bayesian model based on fold change BrdU (Table 2; analysis #7). Lastly, we combined proliferative lesion incidence and BrdU labelling in a Bayesian model, convenient for incorporating multiple data types, to increase certainty in the 2-year carcinogenicity predictions (Table 2; analysis #8).

3.2. Histopathological lesions representative of proliferation after sub-chronic chemical exposure are predictive of 2-year tumour incidence, though with low sensitivity

Response-response curves for sub-chronic proliferative lesion incidence and 2-year tumour incidence were fit for 16 carcinogenic

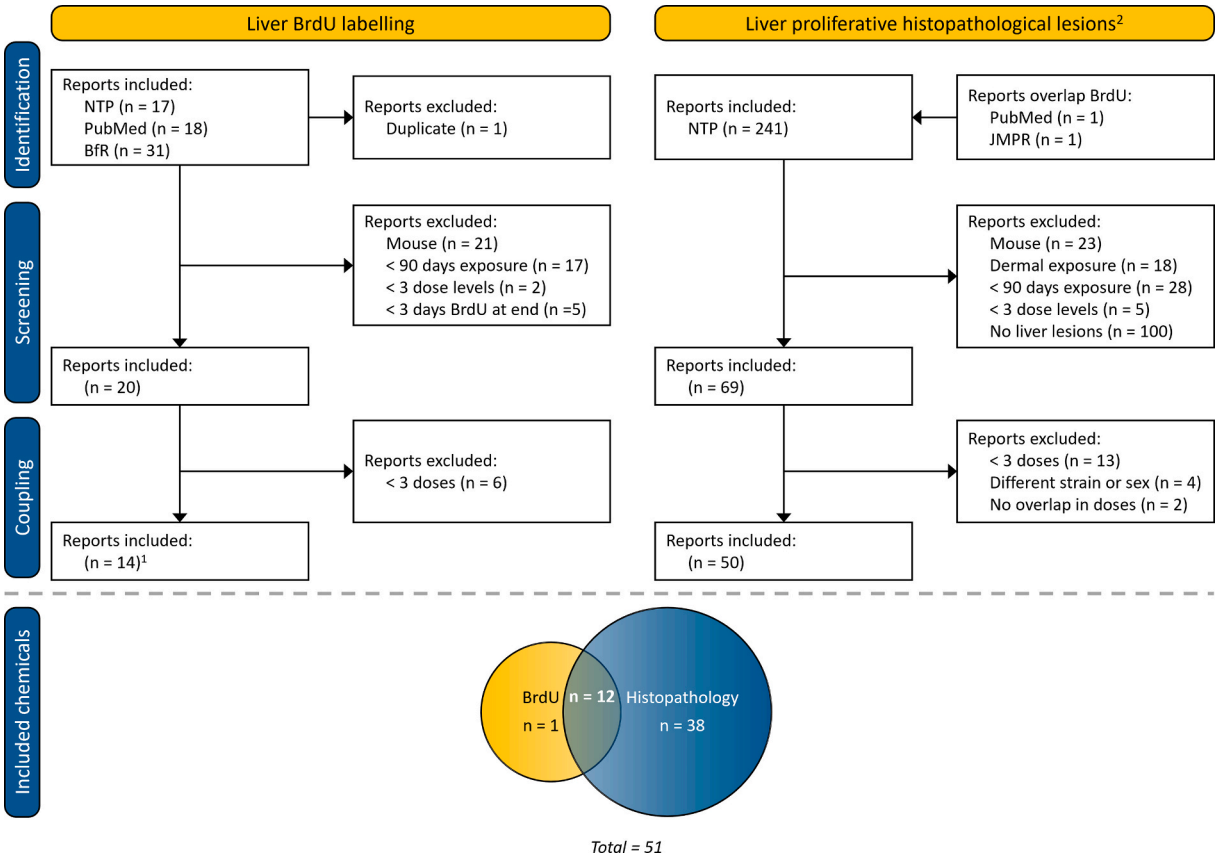


Fig. 3. Data collection. Data from sub-chronic toxicity and 2-year carcinogenicity studies were obtained from existing regulatory reports and scientific publications. "Identification" refers to the initial data search. "Screening" lists the additional criteria set for sub-chronic toxicity studies. Criteria for 2-year carcinogenicity studies are outlined at "Coupling". Numbers represent reports. The number of included chemicals, and their overlap, is depicted at "Included chemicals". [1]A total of 14 reports were included for BrdU, resulting in 13 unique chemicals. [2]Subset of five histopathology lesions: hepatocyte hyperplasia, multinucleated hepatocytes, basophilic foci, acidophilic foci, and mitotic alterations.

Table 2
Overview of performed analyses. Analysis reference number, modelling method (type), predictor (data type) and number of liver carcinogens and liver non-carcinogens used.

Analysis	Method	Predictor	Chemicals
1	Response-Response (effect size)	Extra proliferative lesion incidence (continuous)	Liver carcinogens: 16
2	Bayesian logistic regression (probabilistic)	Presence/absence proliferative lesions (categorical)	Liver carcinogens: 16 Liver non-carcinogens: 34
3	Bayesian logistic regression (probabilistic)	Extra proliferative lesion incidence (continuous)	Liver carcinogens: 16 Liver non-carcinogens: 34
4	Response-Response (effect size)	BrdU fold change over vehicle (continuous)	Liver carcinogens: 10
5	Response-Response (effect size)	BrdU fold change over vehicle (continuous); based on internal liver concentrations	Liver carcinogens: 10
6	Bayesian logistic regression (probabilistic)	Threshold BrdU fold change over vehicle (categorical)	Liver carcinogens: 10 Liver non-carcinogens: 3
7	Bayesian logistic regression (probabilistic)	BrdU fold change over vehicle (continuous)	Liver carcinogens: 10 Liver non-carcinogens: 3
8	Bayesian logistic regression (probabilistic)	Extra proliferative lesion incidence and BrdU fold change over vehicle (combined, continuous)	Liver carcinogens: 10 Liver non-carcinogens: 2

chemicals (supplementary figure S1). As the tested doses between the sub-chronic toxicity and 2-year carcinogenicity study overlapped minimally, response-response curves depend highly on data imputation for tumour incidences. Of the 16 carcinogenic chemicals, 5 could not be fitted using PROAST as there were no proliferative lesions observed in the sub-chronic study (supplementary Table S4). For most chemicals, proliferative lesions of interest were found only in the highest or two highest dose levels tested and observed incidences were often quite high, whereas tumours were observed at lower doses and showed a more gradual increase (supplementary Table S4). As a result, half of the response-response curves showed a very abrupt increase and were thus uninformative, especially for BMD analysis and derivation of a point of departure (PoD) (supplementary figure S1). Likewise, the initial vertical increase in the chemical-agnostic response-response curve indicated that absence of proliferative lesions does not reflect non-carcinogenicity (supplementary figure S2). Nevertheless, the response-response curves did indicate that the presence of proliferative lesions in a sub-chronic study is a relevant predictor of carcinogenicity, as virtually all datapoints with proliferative lesions match with an extra tumour incidence above 10 % (supplementary figure S2).

Given the relatively sparse dataset, we continued with Bayesian logistic regression modelling to predict the probability of carcinogenicity, and its distribution, based on the proliferative lesions observed in sub-chronic toxicity studies. Mean predicted probabilities of 1,000 posterior draws were compared between carcinogenic and non-carcinogenic observations for both a categorical model, based on presence/absence of proliferative lesions, and a continuous model based on incidence of proliferative lesions (Fig. 4A). Few chemical-dose combinations were wrongly predicted as positive, indicating that proliferative lesions are a highly specific predictor. Even so, many were incorrectly predicted as negative, highlighting low sensitivity (Fig. 4B). Most erroneous predictions were made for the lower dose levels tested. As the intended use of these models is to derive a PoD, missing lower doses results in prediction-based BMDs which are higher than the actual outcome-based BMDs. In conclusion, while presence of the selected proliferative lesions has a high positive predictive value, absence is not predictive of non-carcinogenicity, decreasing its regulatory applicability.

3.3. Hepatocyte BrdU labelling after sub-chronic chemical exposure can be used to model an informative response-response relationship between proliferation and tumorigenesis

Besides specific histopathological lesions, BrdU labelling was explored as a marker for proliferation to predict tumorigenesis. Again, response-response curves for sub-chronic BrdU labelling and 2-year tumour incidence were fitted for carcinogenic chemicals (supplementary figure S3). PROAST could not fit a response-response curve for 2 of the 10 carcinogenic chemicals, as these chemicals did not show an increase in BrdU labelling while an increase in tumour incidence was observed. Next, we combined chemical-specific response-response models for the remaining 8 chemicals, to obtain a chemical-agnostic response-response relation with confidence interval (Fig. 5A, blue line). To check if the 2 not-fitted chemicals severely influence the curve, a chemical-agnostic response-response curve was also directly fitted in PROAST using all normalized datapoints (Fig. 5A, yellow line). The resulting curves were quite similar. Importantly, the lower 95 %

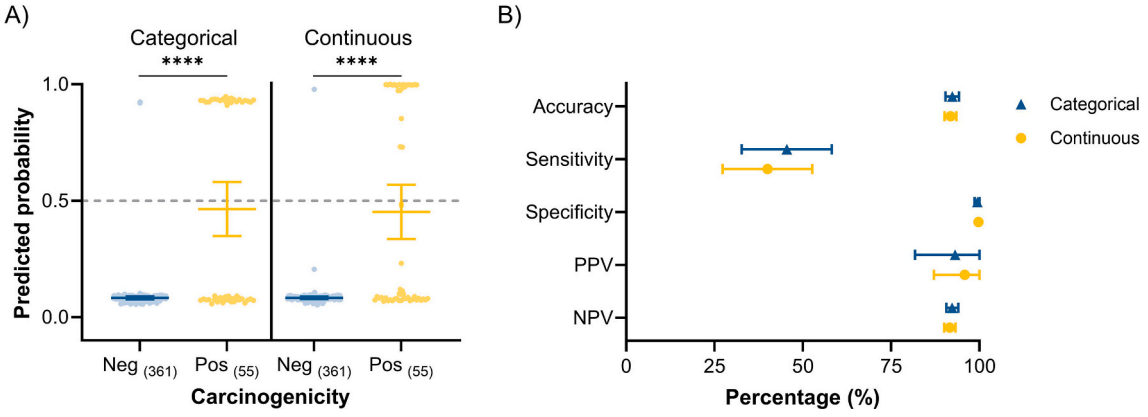


Fig. 4. Proliferative lesions found in sub-chronic studies are predictive of carcinogenicity. A) Mean predicted probabilities of ≥ 10 % extra liver tumour incidence for non-carcinogenic (<10 % extra liver tumour incidence; neg/blue) and carcinogenic (≥10 % extra liver tumour incidence; pos/yellow) chemical-dose combinations determined by 1,000 posterior draws from a categorical and a continuous Bayesian logistic regression model. Numbers in parentheses indicate the total number of datapoints within each group. Whiskers represent the mean and 95 % credible interval. A Mann-Whitney statistical test was used (p-value < 0.0001). B) Measures of diagnostic evaluation for the Bayesian logistic regression models. Dots and error bars represent the mean and 95 % credible interval, respectively. A probability cut-off of 0.5 was used to assign probabilities into positives or negatives. Blue/triangle = categorical model, yellow/circle = continuous model. PPV = positive predictive value, NPV = negative predictive value.

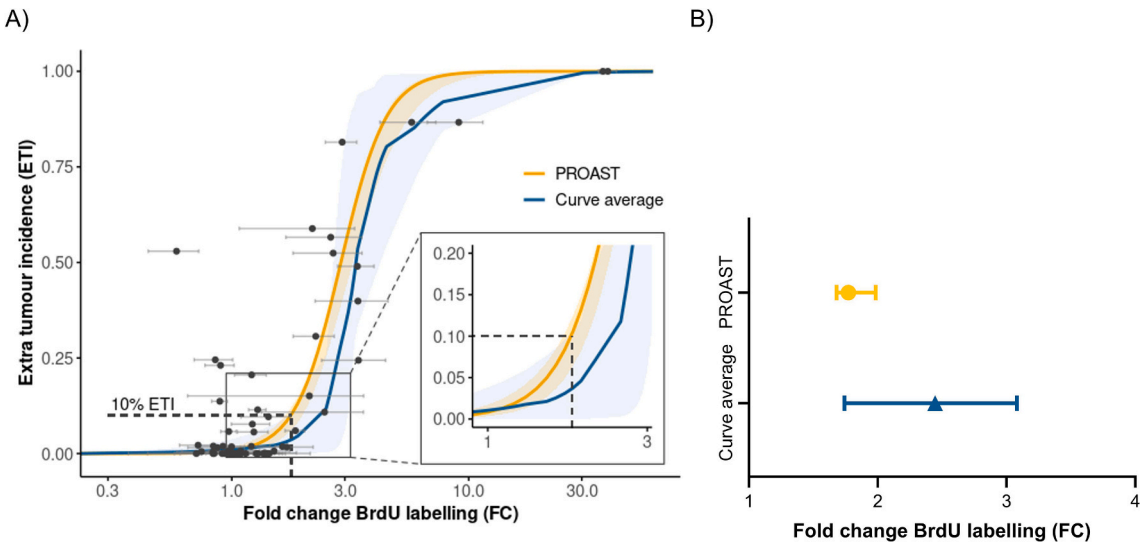


Fig. 5. Response-response curve for 2-year extra liver tumour incidence and fold change BrdU labelling after sub-chronic chemical exposure. A) Chemical-agnostic log-logistic response-response relation including the 95% confidence interval. The blue line results from taking the average of the eight chemical-specific response-response curves. The directly fitted response-response curve by PROAST, based on 10 carcinogenic chemicals, is depicted in yellow. The dotted line represents an extra tumour incidence of 10% and the corresponding fold change BrdU labelling. B) Benchmark dose 10% response with confidence interval for curve average (blue/triangle) and PROAST (yellow/circle) response-response curves.

confidence interval for 10 % extra tumour incidence of the curve average is virtually the same as the one based on the PROAST fit (Fig. 5A enlargement and Fig. 5B). The response-response curve demonstrates that the fold change in BrdU labelling after sub-chronic chemical exposure can be used as a predictor of 2-year extra tumour incidence, and its S-shape is suitable for deriving a PoD.

3.4. Response-response modelling informed by internal dose levels does not significantly alter from administered doses

To validate that the absence of a response was not a result of the chemical not reaching the liver, internal liver concentration data were either collected from the reports or modelled using Simcyp Rat Simulator. All chemicals could either be detected in the liver (through experimental data, as stated in the respective reports), or were predicted to reach the liver using modelling (supplementary Table S6). To account for differences between administered and internal dose levels, response-response modelling was performed for internal liver concentrations (supplementary figure S4-S5). The internal liver concentration-based response-response showed a similar curve as the administered dose-based response-response curve (Fig. 6A), and resulting BMD intervals were largely overlapping (Fig. 6B). Since there were no significant differences between the resulting two curves and BMD intervals, we continued with the response-response model based on administered dose as this model is not based on modelled independent values.

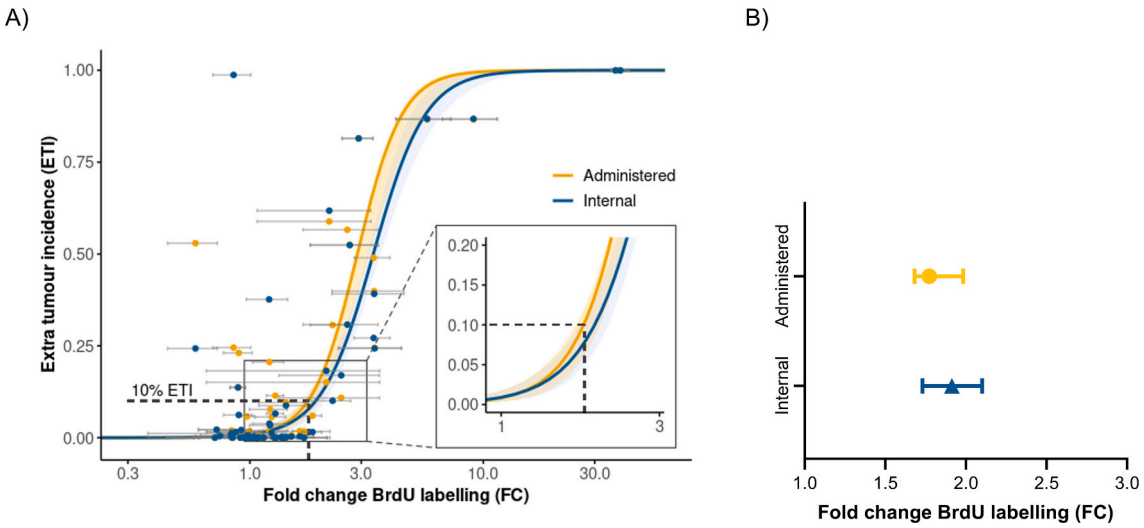


Fig. 6. Response-response curve for 2-year extra liver tumour incidence and fold change BrdU labelling after sub-chronic chemical exposure. A) Chemical-agnostic log-logistic response-response relation based on ten carcinogenic chemicals and the 95% confidence interval. The internal dose–response informed response-response is depicted in blue. The administered dose–response informed response-response curve is depicted in yellow. The dotted line represents an extra tumour incidence of 10% and the corresponding fold change in BrdU labelling. B) Benchmark dose 10% response with confidence interval for internal dose–response (blue/triangle) and administered dose–response (yellow/circle) informed response-response between BrdU fold change and extra tumour incidence.

3.5. Hepatocyte BrdU labelling after sub-chronic chemical exposure is a more sensitive predictor of 2-year tumour incidence compared to proliferative lesions

Next, Bayesian logistic regression models based on the BrdU labelling data were generated to predict the probability of carcinogenicity. We developed both a categorical model based on the threshold obtained from the BrdU fold change response-response curve (1.74) and a continuous model based on fold change BrdU labelling. Mean predicted probabilities of 1,000 posterior draws were compared between carcinogenic and non-carcinogenic data entries (Fig. 7A). In concordance with the response-response model, more, and less pronounced, sub-chronic responses were correctly predicted as carcinogenic, resulting in a higher sensitivity (Fig. 7B). To summarize, fold change BrdU labelling has a slightly lower negative predictive value in comparison to proliferative lesions, yet its higher sensitivity speaks in its favour for deriving accurate PoDs.

3.6. Combining histopathology and BrdU labelling from sub-chronic toxicity studies yields more accurate 2-year tumour carcinogenicity predictions

Next, we combined the proliferative lesions and BrdU labelling in a Bayesian logistic regression model. As the continuous and categorical models for both proliferative lesions and BrdU labelling were minimally different in performance, all four possible combinations were tested and compared. The model consisting of continuous variables showed slightly better performance than the other combinations (supplementary Table S10). Single-variable models of these two predictors on the combined dataset were run for comparison (supplementary Table S10). Once more, mean predicted probabilities of 1,000 posterior draws were compared between carcinogenic and non-carcinogenic chemical-dose combinations (Fig. 8A). We then compared the combined and single-variable prediction models for their accuracy, sensitivity, specificity, PPV and NPV on the combined dataset. The combined model showed considerable improvements over the single-variable models, specifically over the model for proliferative lesion incidence (Fig. 8B). Importantly, for most of the tested chemicals, BMD intervals from the combined model were either more substantially overlapping with, or closer to, the outcome-based BMD intervals compared to the single-variable models (Fig. 8C). Moreover, BMDs could be obtained for more chemicals based on the combined model, whereas for some a single-variable model could not be fit due to absence of the response. Nevertheless, not all prediction-based lower BMD intervals were lower than the 2-year liver tumour lowest observed adverse effect level (LOAEL), indicating that an extra safety factor would be required for regulatory application of this approach. To recapitulate, including fold change BrdU labelling in the prediction modelling greatly improves the overall predictive performance of the model, specifically improving its sensitivity, subsequently allowing for derivation of more accurate PoDs.

3.7. Application of the developed prediction model to obtain a PoD for chemical risk assessment

To exemplify the use of the developed model, we applied the combined Bayesian logistic regression model to two chemicals which were excluded from the original datasets due to insufficient or lack of cancer data: WY-14,643 (WY), a liver carcinogen, and 3,3′,4,4′-tetrachloroazoxybenzene (TCAB), a liver non-carcinogen (supplementary Table S11). For WY, a dose–response curve for predicted probability of carcinogenicity was fit (Fig. 9A, blue line), and a liver cancer BMD was obtained (Fig. 9B, blue). The obtained BMD and its interval were observed within the tested dose range (Fig. 9B, blue). Similarly, a dose–response curve for TCAB predictions was fit (Fig. 9A, yellow line). The probability threshold was not reached for the tested dose range. Therefore, the calculated BMD highly depends on extrapolation and is thus considered non-informative (Fig. 9B, yellow). To conclude, our prediction model labels WY as liver carcinogen and TCAB as liver non-carcinogen, concordant with their respective reports, and PoDs for risk assessment were obtained.

4. Discussion

We reported the quantification of the KER between sustained proliferation and tumour induction in rat liver using different modelling approaches. We illustrated that response-response modelling is minimally affected by differences between administered doses and internal liver concentrations. Additionally, we showed that combined modelling of proliferative lesion incidence and BrdU labelling allows for more accurate predictions of tumour formation compared to predictions based solely on proliferative lesion incidence. Importantly, our modelling approach allows for determination of a PoD based on 90-day sub-chronic

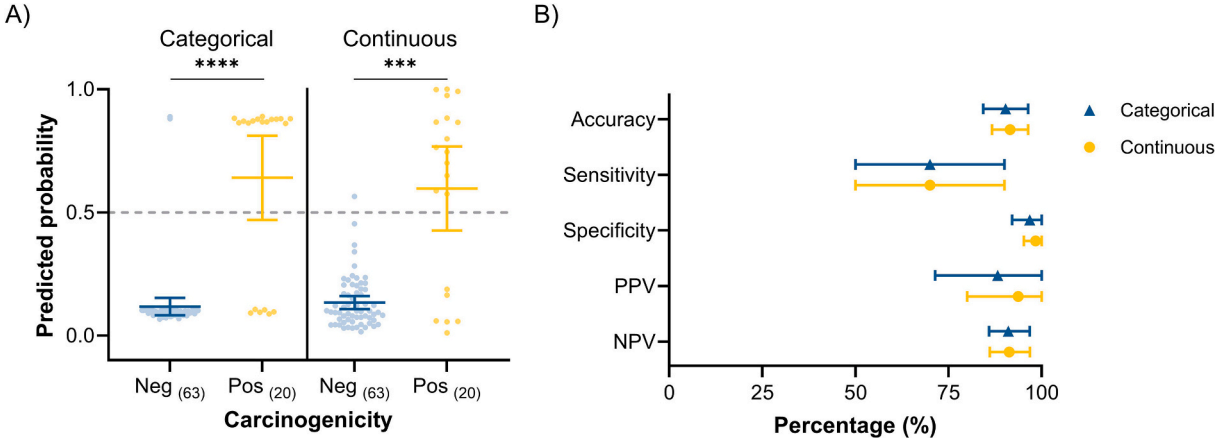
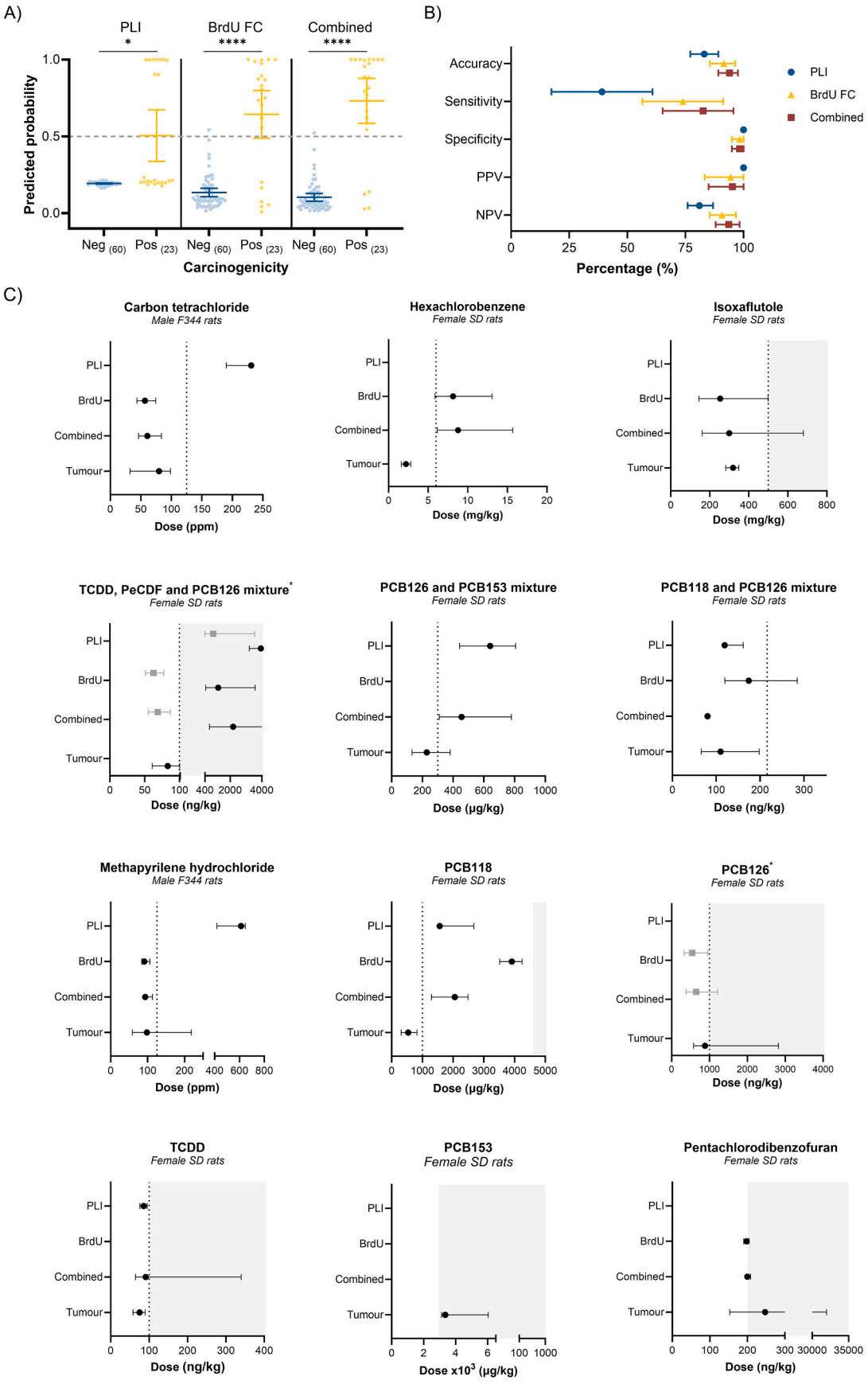


Fig. 7. BrdU labelling after sub-chronic chemical exposure is predictive of 2-year tumour incidence. A) Mean predicted probabilities of ≥ 10 % extra liver tumour incidence for non-carcinogenic (<10 % extra liver tumour incidence; neg/blue) and carcinogenic (≥10 % extra liver tumour incidence; pos/yellow) chemical-dose combinations by 1,000 posterior draws from a categorical and a continuous Bayesian logistic regression model. Numbers in parentheses indicate the total number of datapoints within each group. Whiskers represent the mean and 95 % credible interval. A Mann-Whitney statistical test was used (p-value **** < 0.0001, *** < 0.0005). B) Measures of diagnostic evaluation for the Bayesian logistic regression model. Dots represent the mean and errors bars the 95 % credible interval. A probability cut-off of 0.5 was used to assign probabilities into positives or negatives. Blue/triangle = categorical model, yellow/circle = continuous model. PPV = positive predictive value, NPV = negative predictive value.



(caption on next page)

Fig. 8. 2-year extra tumour incidence probability predictions based on histopathological lesions and BrdU labelling. A) Mean predicted probabilities of ≥ 10 % extra liver tumour incidence for non-carcinogenic (<10 % extra liver tumour incidence; neg/blue) and carcinogenic (≥10 % extra liver tumour incidence; pos/yellow) chemical-dose combinations by 1,000 posterior draws from the three Bayesian logistic regression models. Numbers in parentheses indicate the total number of datapoints within each group. Whiskers represent the 95 % credible interval. A Mann-Whitney statistical test was used (p-value * < 0.05; **** < 0.0001). PLI = proliferative lesion incidence, BrdU FC = fold change BrdU labelling, Combined = proliferative lesion incidence and fold change BrdU labelling. B) Measures of diagnostic evaluation for the Bayesian logistic regression models based on 1,000 draws from the posterior distribution. Dots represent the mean and errors bars the 95 % credible interval. A probability cut-off of 0.5 was used. Blue/circle = proliferative lesion incidence, yellow/triangle = fold change BrdU labelling, red/square = combined. PLI = proliferative lesion incidence, BrdU FC = fold change BrdU labelling, Combined = proliferative lesion incidence and fold change BrdU labelling. PPV = positive predictive value, NPV = negative predictive value. C) Prediction-based benchmark doses for 10 % extra tumour incidence versus outcome-based benchmark doses for 10 % extra tumour incidence for 10 liver carcinogens and 2 liver non-carcinogens. Dots represent the mean and whiskers the 95 % credible interval. Grey area represents doses outside the tested range, and the dotted line represents the 2-year liver tumour lowest observed adverse effect level (LOAEL). *Gray squares represent benchmark doses calculated excluding the highest tested dose, as it deviated from the expected dose–response pattern.

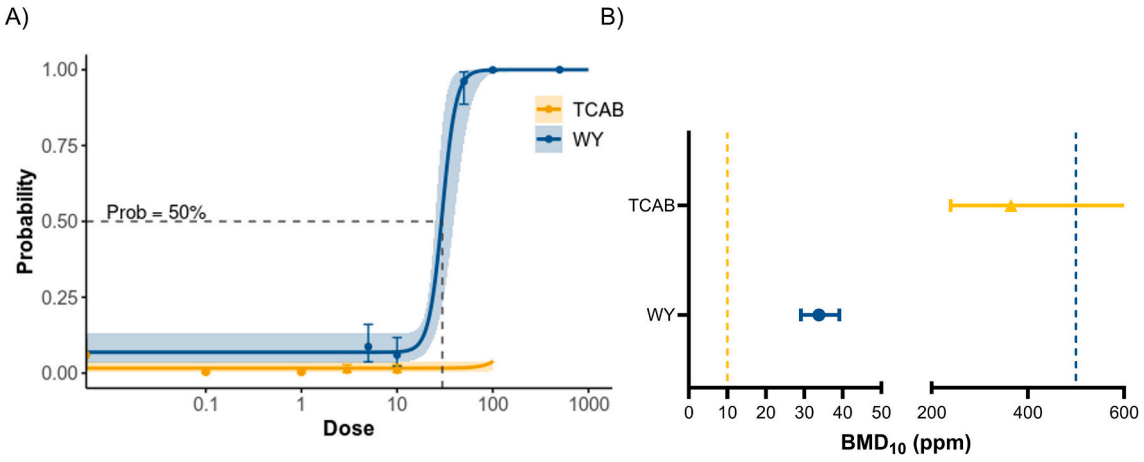


Fig. 9. Predicted dose–response curves and benchmark dose intervals for a liver carcinogen and a non-carcinogen. A) Dose-responses of mean predicted probability of carcinogenicity for WY (blue; liver carcinogen) and TCAB (yellow; liver non-carcinogen). The 95% credible interval is highlighted in the respective colour. The dotted line represents a 50% probability of carcinogenicity. B) Benchmark dose with confidence interval for a 50% probability of carcinogenicity for WY (blue; liver carcinogen) and TCAB (yellow; liver non-carcinogen). The dotted line in the respective colour highlights the maximum tested dose.

toxicity data that is close to the 2-year cancer bioassay tumour incidence PoD.

The systematically compiled dataset includes data from 60 reports on histopathological proliferative lesion incidence, BrdU labelling and liver tumour incidence in rats. We obtained a harmonized dataset by following predefined criteria (Fig. 3), despite combining data originating from various laboratories, in a period of decades, using different strains and various exposure routes. Importantly, more than half of the data consists of non-carcinogenic dose levels (approximately 72 % in the combined dataset). By performing Bayesian logistic regression and incorporation of priors, we accounted for prediction bias associated with the scarceness of carcinogenic data entries.

An important limitation when interpretating the BrdU labelling index is that it reflects the proportion of replicating cells and does not account for a potential increase in the absolute number of cells, such as the result of hyperplasia, which may lead to an underestimation of the overall proliferative effect. Since the liver has a relatively slow proliferation rate, resulting in low baseline BrdU labelling, we opted to use the fold change to express differences between treatment conditions and the vehicle control. As this inherently normalizes differences relative to the baseline, it can more effectively highlight small changes, which are expected to be biologically relevant given the slow proliferation rate of the liver. BrdU labelling is determined by counting the number of labelled cells in at least 2,000 hepatocellular nuclei, often more than once per animal. Nevertheless, there is a high sensitivity to incidental findings. Therefore, we applied a threshold to the BrdU fold change to mitigate the impact of outliers and minor fluctuations (categorical BrdU fold change model; Table 2, analysis #6). We selected this threshold using the response-response model for the relationship between BrdU labelling and tumour incidence, and applied this in Bayesian logistic regression analysis. Our selected threshold, 1.74-fold, aligns with commonly used thresholds of 1.5- to 2-fold differences. Of the six datapoints with an extra tumour incidence ≥ 10 % that did not reach the threshold in BrdU labelling, five were from reports which also included BrdU labelling after 31 and 53 weeks of chemical exposure [74–78]. Four out of these five chemical-dose combinations did show a difference in BrdU labelling > 1.74-fold after 31 weeks of chemical exposure. For these particular chemical-dose combinations it seems that 90-days is too early to observe a proliferative effect in the liver as a consequence of the chemical exposure. Expanding this dataset by collecting BrdU labelling data for additional chemicals could provide insight into whether this effect is related to chemical properties or represents a limitation inherent to the use of 90-day data.

A preliminary list of potential sources of uncertainty, and their qualitative impact on the model's outcome were assessed (Table 3). Three of these will be discussed in more depth. To start, credibility of a quantitative AOP can never exceed the confidence in the underlying qualitative AOP [31]. The causal link underlying the relationship between proliferation and tumour induction is well-established [34,35,33], although not all cases of sustained proliferation proceed to tumour formation, requiring abnormal signalling or impaired homeostasis as well [79,80]. Furthermore, carcinogenesis is a stochastic process, inherently involving an element of chance. Nevertheless, the AOP containing this KER [48] has an endorsed status, reflecting thorough review of the causal links and the uncertainty impact on the predictions is thus expected to be minimal. Secondly, *in vivo* data is inherently associated with uncertainty due to biological diversity and experimental design. Biological variability arises from both genetic and physiological differences, such as sex and hormone cycles [81,82,83,84]. Uncertainty as a consequence of experimental design mainly includes heterogeneity in administered doses [85,86] and environmental factors such as housing, temperature and type of diet [87,84]. Lastly, certain measurement

Table 3
Potential sources of uncertainty and their expected impact on the model's outcome.

Source of uncertainty	Impact	Reasoning
AOP structure	Low	The causal link was inferred from a well-established relation.
In vivo data	High	*In vivo* data are associated with large intra-chemical variability and uncertainty resulting from, amongst others, individual variation, dosing method, species- and strain-effects.
Experimental data	High	Studies were performed over a time course of 30 years, with different exposure routes, strains and sexes being used.
Reuse of data	Moderate	As the dataset was not specifically designed for computational approaches, it is not ideally suited for modelling purposes (e.g., modelling would benefit from all readouts being reported in individuals instead of group averages).
Data imputation	High	To maximize overlap between the two KEs, tumour incidences were obtained through data imputation based on fitted dose–response curves. Imputed data was then used for response-response modelling and Bayesian logistic regression analysis.
Modelling method	Low	A major challenge in qAOP development is choosing the right model for the data. Both the predictors and outcome were considered quantal in their nature, and probabilistic modelling is well-suited for quantal outcomes. By developing both a probabilistic and effect size prediction model, we could model both the distribution and magnitude of outcomes.
Choice of priors	Low	We tried to minimize the effect of the priors on the posterior distribution by using weakly informative priors.
Model robustness	Low	Model parameters showed the (Bayesian) models converged well.
Model output	Low	Mean and credible/confidence intervals are informative.
Chemical independence	Moderate	The training dataset contains chemicals from different domains (pharmaceuticals, agrochemicals and industrial chemicals). No selection on physicochemical properties, nor mode of action has been made. Nevertheless, we do realize that our models are trained on a limited dataset.
Applicability domain	High	Application of the models is currently limited to predicting chronic rat liver effects based on sub-chronic rat liver effects.

techniques, such as histopathological scoring, are subjective to human interpretation bias [88], which can be partially mitigated through blinded assessment. Taken together, we scored the uncertainty impact of using *in vivo* data as high. Thirdly, the applicability domain of the model is currently limited to 2-year rat liver predictions based on 90-day rat liver data. It would be valuable to explore the potential for making 2-year predictions based on 28-day data, or even shorter exposure durations, and for other species and organs. However, we found the currently available data for such analysis, that suffices to our data selection criteria, is insufficient. In other words, predictions for other exposure durations, species and organs cannot reliably be made with the current model. Notably, a time conversion factor is typically used in regulatory toxicology to correct for the difference in exposure duration between chronic and sub-chronic toxicity studies [89]. Yet, as we modelled the relation between sub-chronic biomarkers and a chronic effect, time is inherently incorporated into the model. Therefore, within the context of our approach, there is no need to incorporate the time conversion factor into the current safety factors.

With the current shift in regulatory toxicology to human-centred, mechanism-based risk assessment, there is a need for quantitative AOPs to link short term *in vitro* effects to long term *in vivo* adverse outcomes. Moreover, including earlier KEs, that can be observed in *in vitro* test systems, strengthens the mechanistic insights of qAOP modelling. For bridging early KE *in vitro* information to late KEs that are observed *in vivo* (i.e. histopathology-based proliferation markers and BrdU labelling) PBK-modelling and quantitative *in vitro* to *in vivo* extrapolation are required. Additionally, PBK-modelling can be used for species extrapolation (e.g., [90]) and modelling of various or aggregated exposures (e.g., [91]).

The current models are developed on a limited set of chemicals due to data availability. Therefore, to further explore this quantitative KER and its regulatory application, it is our recommendation to standardize and include BrdU labelling in sub-chronic toxicity studies for regulatory purposes. Despite the extra practises, a harmonized dataset covering the chemical domain is needed for model improvement, tailoring to specific chemical properties and validation, including the establishment of historical control data. As a critical effect size of 10 % is the most commonly used threshold for quantal data in a regulatory context [53], we applied this 10 % threshold in our analyses. Nevertheless, the applied modelling method allows for customization of the threshold to specific regulatory needs or questions. Implementation of this quantitative KER allows for carcinogenic risk assessment without the need for a 2-year carcinogenicity study, contributing to reduction of animal experimentation by combining sub-chronic toxicity data with the demonstrated modelling approach. Importantly, the methods applied, in conjunction with PBK-modelling, can be further extended to base predictions on earlier KEs quantified using *in vitro* data, ultimately aiding the transition to animal-free chemical risk assessment.

CRediT authorship contribution statement

Christina H.J. Veltman: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Hiba Khalidi:** Writing – review & editing, Formal analysis. **Elias Zgheib:** Writing – review & editing, Formal analysis. **Bob van de Water:** Writing – review & editing, Supervision. **Mirjam Luijten:** Writing – review & editing, Supervision, Conceptualization. **Jeroen L.A. Pennings:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Dr. Ruud Woutersen for sharing his expertise in histopathology and advice on the selection of proliferative lesions; Dr. André Muller for sharing his expertise in regulatory toxicology and advice on commonly applied thresholds in hazard assessment; and Mark Cronin and Dr. Véronique de Bruijn for carefully reviewing the manuscript.

Funding sources

This project received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No. 964537 (RISK-HUNT3R), and it is part of the ASPIS cluster.

Disclaimer

The opinions expressed in this document reflect only the author's views. The European Commission is not responsible for any use that may be made of the information it contains.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.comtox.2025.100359.

Data availability

Data were collected from public reports, with references and datasets provided in the supplementary files.

References

[1] D. Hanahan, R.A. Weinberg, The hallmarks of cancer, *Cell* 100 (1) (2000) 57–70, https://doi.org/10.1016/s0092-8674(00)81683-9.

[2] H. Heusinkveld, H. Braakhuis, R. Gommans, P. Botham, M. Corvaro, J.W. van der Laan, D. Lewis, F. Madia, I. Manou, F. Schorsch, G. Wolterink, R. Woutersen, R. Corvi, J. Mehta, M. Luijten, Towards a mechanism-based approach for the prediction of nongenotoxic carcinogenic potential of agrochemicals, *Crit Rev Toxicol* 50 (9) (2020) 725–739, https://doi.org/10.1080/10408444.2020.1841732.

[3] M. Luijten, R. Corvi, J. Mehta, M. Corvaro, N. Delrue, S. Felter, B. Haas, N. J. Hewitt, G. Hilton, T. Holmes, M.N. Jacobs, A. Jacobs, F. Lamplmair, D. Lewis, F. Madia, I. Manou, S. Melching-Kollmuss, F. Schorsch, K. Schutte, H. Braakhuis, A comprehensive view on mechanistic approaches for cancer risk assessment of non-genotoxic agrochemicals, Regul Toxicol Pharmacol 118 (2020), https://doi.org/10.1016/j.yrtph.2020.104789.

[4] S.M. Cohen, Cell proliferation and carcinogenesis, *Drug Metab Rev* 30 (2) (1998) 339–357, https://doi.org/10.3109/03602539808996317.

[5] M.N. Jacobs, A. Colacci, R. Corvi, M. Vaccari, M.C. Aguila, M. Corvaro, N. Delrue, D. Desaulniers, N. Ertych, A. Jacobs, M. Luijten, F. Madia, A. Nishikawa, K. Ogawa, K. Ohmori, M. Paparella, A.K. Sharma, P. Vasseur, Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens, *Arch Toxicol* 94 (8) (2020) 2899–2923, https://doi.org/10.1007/s00204-020-02784-5.

[6] OECD, *Test No. 451: Carcinogenicity Studies*, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 2018. https://doi.org/10.1787/9789264071186-en.

[7] OECD. *Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies*, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 2018. https://doi.org/10.1787/9789264071223-en.

[8] E. Gottmann, S. Kramer, B. Pfahringer, C. Helma, Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments, Environ Health Perspect 109 (5) (2001) 509–514, https://doi.org/10.1289/ehp.01109509.

[9] P.A. Marone, W.C. Hall, A.W. Hayes, Reassessing the two-year rodent carcinogenicity bioassay: a review of the applicability to human risk and current perspectives, *Regul Toxicol Pharmacol* 68 (1) (2014) 108–118, https://doi.org/10.1016/j.yrtph.2013.11.011.

[10] M. Luijten, E.D. Olthof, B.C. Hakkert, E. Rorije, J.W. van der Laan, R.A. Woutersen, J. van Benthem, An integrative test strategy for cancer hazard identification, *Crit Rev Toxicol* 46 (7) (2016) 615–639, https://doi.org/10.3109/10408444.2016.1171294.

[11] L.G. Hernandez, H. van Steeg, M. Luijten, J. van Benthem, Mechanisms of non-genotoxic carcinogens and importance of a weight of evidence approach, *Mutat Res* 682 (2–3) (2009) 94–109, https://doi.org/10.1016/j.mrrev.2009.07.002.

[12] OECD. (2018). Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways. *OECD Series on Adverse Outcome Pathways, No. 1*. https://doi.org/10.1787/5jlv1m9d1g32-en.

[13] D.L. Villeneuve, D. Crump, N. Garcia-Reyero, M. Hecker, T.H. Hutchinson, C. A. LaLone, B. Landesmann, T. Lettieri, S. Munn, M. Nepelska, M.A. Ottinger, L. Vergauwen, M. Whelan, Adverse outcome pathway (AOP) development I: strategies and principles, *Toxicol Sci* 142 (2) (2014) 312–320, https://doi.org/10.1093/toxsci/kfu199.

[14] M.N. Jacobs, A. Colacci, K. Louekari, M. Luijten, B.C. Hakkert, M. Paparella, P. Vasseur, International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances, *ALTEX* 33 (4) (2016) 359–392. https://doi.org/10.14573/altex.1601201.

[15] S. Schmeisser, A. Miccoli, M. von Bergen, E. Berggren, A. Braeuning, W. Busch, C. Desaintes, A. Gourmelon, R. Grafstrom, J. Harrill, T. Hartung, M. Herzler, G.E. N. Kass, N. Kleinstreuer, M. Leist, M. Luijten, P. Marx-Stoelting, O. Poetz, B. van Ravenzwaay, T. Tralau, New approach methodologies in human regulatory toxicology - Not if, but how and when!, Environ Int 178 (2023), https://doi.org/10.1016/j.envint.2023.108082.

[16] C.A. LaLone, G.T. Ankley, S.E. Belanger, M.R. Embry, G. Hodges, D. Knapen, S. Munn, E.J. Perkins, M.A. Rudd, D.L. Villeneuve, M. Whelan, C. Willett, X. Zhang, M. Hecker, Advancing the adverse outcome pathway framework-An international horizon scanning approach, *Environ Toxicol Chem* 36 (6) (2017) 1411–1421, https://doi.org/10.1002/etc.3805.

[17] E.J. Perkins, R. Ashauer, L. Burgoon, R. Conolly, B. Landesmann, C. Mackay, C. A. Murphy, N. Pollesch, J.R. Wheeler, A. Zupanic, S. Scholz, Building and Applying Quantitative Adverse Outcome Pathway Models for Chemical Hazard and Risk Assessment, *Environ Toxicol Chem* 38 (9) (2019) 1850–1865, https://doi.org/10.1002/etc.4505.

[18] L. Bajard, O. Adamovsky, K. Audouze, K. Baken, R. Barouki, J.B. Beltman, A. Beronius, E.C. Bonefeld-Jørgensen, G. Cano-Sancho, de Baat, F. Di Tillio, M. F. Fernández, R.E. FitzGerald, C. Gundacker, A.F. Hernández, K. Hilscherova, S. Karakitsios, E. Kuchovska, M. Long, L. Blaha, Application of AOPs to assist regulatory assessment of chemical risks - Case studies, needs and recommendations, Environ Res 217 (2023), https://doi.org/10.1016/j.envres.2022.114650.

[19] R.B. Conolly, G.T. Ankley, W. Cheng, M.L. Mayo, D.H. Miller, E.J. Perkins, D. L. Villeneuve, K.H. Watanabe, Quantitative Adverse Outcome Pathways and Their Application to Predictive Toxicology, *Environ Sci Technol* 51 (8) (2017) 4661–4672, https://doi.org/10.1021/acs.est.6b06230.

[20] J.A. Doering, D.L. Villeneuve, S.T. Poole, B.R. Blackwell, K.M. Jensen, M.D. Kahl, A.R. Kittelson, D.J. Feifarek, C.B. Tilton, C.A. LaLone, G.T. Ankley, Quantitative Response-Response Relationships Linking Aromatase Inhibition to Decreased Fecundity are Conserved Across Three Fishes with Asynchronous Oocyte Development, *Environ Sci Technol* 53 (17) (2019) 10470–10478, https://doi.org/10.1021/acs.est.9b02606.

[21] C.M. Foran, T. Rycroft, J. Keisler, E.J. Perkins, I. Linkov, N. Garcia-Reyero, A modular approach for assembly of quantitative adverse outcome pathways, *ALTEX* 36 (3) (2019) 353–362. https://doi.org/10.14573/altex.1810181.

[22] I. Hassan, H. El-Masri, P.A. Kosian, J. Ford, S.J. Degitz, M.E. Gilbert, Neurodevelopment and Thyroid Hormone Synthesis Inhibition in the Rat: Quantitative Understanding Within the Adverse Outcome Pathway Framework, *Toxicol Sci* 160 (1) (2017) 57–73, https://doi.org/10.1093/toxsci/kfx163.

[23] S. Jarzina, S. Di Fiore, B. Ellinger, P. Reiser, S. Frank, M. Glaser, J. Wu, F. J. Taverne, N.I. Kramer, A. Mally, Application of the Adverse Outcome Pathway Concept to In Vitro Nephrotoxicity Assessment: Kidney Injury due to Receptor-Mediated Endocytosis and Lysosomal Overload as a Case Study, *Front Toxicol* 4 (2022), https://doi.org/10.3389/ftox.2022.864441.

[24] C. Tebby, W. Gao, J. Delp, G. Carta, W. van der Stel, M. Leist, P. Jennings, B. van de Water, F.Y. Bois, A quantitative AOP of mitochondrial toxicity based on data from three cell lines, *Toxicol in Vitro* 81 (2022), https://doi.org/10.1016/j.tiv.2022.105345.

[25] L.D. Burgoon, M. Angrish, N. Garcia-Reyero, N. Pollesch, A. Zupanic, E. Perkins, Predicting the Probability that a Chemical Causes Steatosis Using Adverse Outcome Pathway Bayesian Networks (AOPBNs), *Risk Anal* 40 (3) (2020) 512–523, https://doi.org/10.1111/risa.13423.

[26] J.S. Jaworska, A. Natsch, C. Ryan, J. Strickland, T. Ashikaga, M. Miyazawa, Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy, *Arch Toxicol* 89 (12) (2015) 2355–2383, https://doi.org/10.1007/s00204-015-1634-2.

[27] J. Jeong, J. Choi, Quantitative adverse outcome pathway (qAOP) using bayesian network model on comparative toxicity of multi-walled carbon nanotubes (MWCNTs): safe-by-design approach, *Nanotoxicology* 16 (5) (2022) 679–694, https://doi.org/10.1080/17435390.2022.2140615.

[28] S.J. Moe, R. Wolf, L. Xie, W.G. Landis, N. Kotamäki, K.E. Tollefsen, Quantification of an Adverse Outcome Pathway Network by Bayesian Regression and Bayesian Network Modeling, *Integr Environ Assess Manag* 17 (1) (2021) 147–164, https://doi.org/10.1002/ieam.4348.

[29] N. Spînu, M.T.D. Cronin, J. Lao, A. Bal-Price, I. Campia, S.J. Enoch, J.C. Madden, L. Mora Lagares, M. Novič, D. Pamies, S. Scholz, D.L. Villeneuve, A.P. Worth, Probabilistic modelling of developmental neurotoxicity based on a simplified adverse outcome pathway network, *Comput Toxicol* 21 (2022), https://doi.org/10.1016/j.comtox.2021.100206.

[30] E. Zgheib, W. Gao, A. Limonciel, H. Aladjov, H. Yang, C. Tebby, G. Gayraud, P. Jennings, M. Sachana, J.B. Beltman, F.Y. Bois, Application of three approaches for quantitative AOP development to renal toxicity, *Comput Toxicol* 11 (2019) 1–13, https://doi.org/10.1016/j.comtox.2019.02.001.

[31] A. Paini, I. Campia, M.T.D. Cronin, D. Asturiol, L. Ceriani, T.E. Exner, W. Gao, C. Gomes, J. Kruisselbrink, M. Martens, M.E.B. Meek, D. Pamies, J. Pletz, S. Scholz, A. Schüttler, N. Spînu, D.L. Villeneuve, C. Wittwehr, A. Worth, M. Luijten, Towards a qAOP framework for predictive toxicology - Linking data to decisions, *Comput Toxicol* 21 (2022), https://doi.org/10.1016/j.comtox.2021.100195.

[32] Society for Advancement of AOPs. *AOP-Wiki*. Retrieved July 18 from https://aopwiki.org/.

[33] C.H.J. Veltman, J.L.A. Pennings, B. van de Water, M. Luijten, An Adverse Outcome Pathway Network for Chemically Induced Oxidative Stress Leading to (Non) genotoxic Carcinogenesis, *Chem Res Toxicol* 36 (6) (2023) 805–817, https://doi.org/10.1021/acs.chemrestox.2c00396.

[34] S.M. Cohen, L.B. Ellwein, Genetic errors, cell proliferation, and carcinogenesis, *Cancer Res* 51 (24) (1991) 6493–6505. https://www.ncbi.nlm.nih.gov/pubmed/1742722.

[35] C. Strupp, M. Corvaro, S.M. Cohen, J.C. Corton, K. Ogawa, L. Richert, M.N. Jacobs, Increased Cell Proliferation as a Key Event in Chemical Carcinogenesis: Application in an Integrated Approach for the Testing and Assessment of Non-Genotoxic Carcinogenesis, *Int J Mol Sci* 24 (17) (2023), https://doi.org/10.3390/ijms241713246.

[36] D.G. Allen, G. Pearse, J.K. Haseman, R.R. Maronpot, Prediction of rodent carcinogenesis: an evaluation of prechronic liver lesions as forecasters of liver tumors in NTP carcinogenicity studies, *Toxicol Pathol* 32 (4) (2004) 393–401, https://doi.org/10.1080/01926230490440934.

[37] A.R. Boobis, S.M. Cohen, N.G. Doerrer, S.M. Galloway, P.J. Haley, G.C. Hard, F. G. Hess, J.S. Macdonald, S. Thibault, D.C. Wolf, J. Wright, A data-based assessment

of alternative strategies for identification of potential human cancer hazards, *Toxicol Pathol* 37 (6) (2009) 714–732, https://doi.org/10.1177/0192623309343779.
[38] C.R. Elcombe, J. Odum, J.R. Foster, S. Stone, S. Hasmall, A.R. Soames, I. Kimber, J. Ashby, Prediction of rodent nongenotoxic carcinogenesis: evaluation of biochemical and tissue changes in rodents following exposure to nine nongenotoxic NTP carcinogens, *Environ Health Perspect* 110 (4) (2002) 363–375, https://doi.org/10.1289/ehp.02110363.
[39] A. Jacobs, Prediction of 2-year carcinogenicity study results for pharmaceutical products: how are we doing? *Toxicol Sci* 88 (1) (2005) 18–23, https://doi.org/10.1093/toxsci/kfi248.
[40] M.V. Reddy, F.D. Sistare, J.S. Christensen, J.G. Deluca, G.K. Wollenberg, J. J. Degeorge, An evaluation of chronic 6- and 12-month rat toxicology studies as predictors of 2-year tumor outcome, *Vet Pathol* 47 (4) (2010) 614–629, https://doi.org/10.1177/0300985810373242.
[41] F.D. Sistare, D. Morton, C. Alden, J. Christensen, D. Keller, S.D. Jonghe, R. D. Storer, M.V. Reddy, A. Kraynak, B. Trela, J.G. Bienvenu, S. Bjurstrom, V. Bosmans, D. Brewster, K. Colman, M. Dominick, J. Evans, J.R. Hailey, L. Kinter, J.J. DeGeorge, An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: support for a proposal to modify current regulatory guidelines, Toxicol Pathol 39 (4) (2011) 716–744, https://doi.org/10.1177/0192623311406935.
[42] J.W. van der Laan, W.H. Buitenhuis, L. Wagenaar, A.E. Soffers, E.P. van Someren, C.A. Krul, R.A. Woutersen, Prediction of the Carcinogenic Potential of Human Pharmaceuticals Using Repeated Dose Toxicity Data and Their Pharmacological Properties, *Front Med* 3 (2016), https://doi.org/10.3389/fmed.2016.00045.
[43] R.A. Woutersen, A.E. Soffers, E.D. Kroese, C.A. Krul, J.W. van der Laan, J. van Benthem, M. Luijten, Prediction of carcinogenic potential of chemicals using repeated-dose (13-week) toxicity data, *Regul Toxicol Pharmacol* 81 (2016) 242–249, https://doi.org/10.1016/j.yrtph.2016.09.003.
[44] A.S. Tischler, J.F. Powers, M. Pignatello, P. Tsokas, J.C. Downing, R.M. McClain, Vitamin D3-induced proliferative lesions in the rat adrenal medulla, *Toxicol Sci* 51 (1) (1999) 9–18, https://doi.org/10.1093/toxsci/51.1.9.
[45] V.F.H. Jensen, P.R. Brinck, J. Nowak, I. Thorup, I. Sjogren, A.M. Molck, Evaluation of Cell Proliferation in Rat Mammary Glands Is Not Predictive of the Carcinogenic Potential of Insulin In Vivo, *Int J Toxicol* 39 (6) (2020) 560–576, https://doi.org/10.1177/1091581820941776.
[46] K. Kawamoto, K. Ogata, H. Asano, K. Miyata, T. Sukata, T. Utsumi, S.M. Cohen, T. Yamada, Cell proliferation analysis is a reliable predictor of lack of carcinogenicity: Case study using the pyrethroid imiprothrin on lung tumorigenesis in mice, *Regul Toxicol Pharmacol* 113 (2020), https://doi.org/10.1016/j.yrtph.2020.104646.
[47] Y. Uno, K. Matsuura, M. Miyagawa, H. Takasawa, H. Tanifuji, K. Abe, A. Akimoto, K. Asanoma, K. Baba, H. Daigo, T. Hagiwara, K. Hirano, T. Inoue, Y. Kawano, K. Kijima, F. Sato, K. Shibata-Yoshida, E. Yamamura, Rat liver in vivo replicative DNA synthesis test for short-term prediction of nongenotoxic (Ames-negative) hepatocarcinogenicity: a collaborative study of the Nongenotoxic Carcinogen Study Group of Japan, *Toxicol Lett* 109 (1–2) (1999) 105–114, https://doi.org/10.1016/s0378-4274(99)00129-0.
[48] Webster, F., Lambert, I. B., & Yauk, C. L. (2017, 29/04/2023). *Cyp2E1 activation leading to liver cancer*. Society for the Advancement of Adverse Outcome Pathways. Retrieved December 6 from https://aopwiki.org/aops/220.
[49] OECD, Test No. 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents. OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 2018. https://doi.org/10.1787/9789264070707-en.
[50] OECD, Test No. 413: Subchronic Inhalation Toxicity: 90-day Study. OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 2018. https://doi.org/10.1787/9789264070806-en.
[51] National Institute for Public Health and the Environment (RIVM), PROAST (2012) Modification date 09/04/2024. Retrieved January 12 from https://www.rivm.nl/en/proast.
[52] W. Slob, Dose-response modeling of continuous endpoints, Toxicol Sci 66 (2) (2002) 298–312, https://doi.org/10.1093/toxsci/66.2.298.
[53] H. Ocagli, R. Zgheib, D. Gregori, S. Belluco, M. Dacasto, I. Baldi, Inventory of BMR values for BMD analysis, *EFSA Supporting Publ.* 21 (4) (2024), https://doi.org/10.2903/sp.efsa.2024.EN-8761.
[54] P.-C. Bürkner, brms: An R Package for Bayesian Multilevel Models Using Stan, *J Stat Softw* 80 (1) (2017). https://doi.org/10.18637/jss.v080.i01.
[55] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, M. Muller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinf.* 12 (77) (2011), https://doi.org/10.1186/1471-2105-12-77.
[56] A. Vehtari, J. Gabry, M. Magnusson, Y. Yao, P.-C. Bürkner, T. Paananen, A. Gelman, (03/07/2024). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models, 2020. https://mc-stan.org/loo/.
[57] A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, *Stat Comput* 27 (5) (2016) 1413–1432, https://doi.org/10.1007/s11222-016-9696-4.
[58] M. Jamei, S. Marciniak, K. Feng, A. Barnett, G. Tucker, A. Rostami-Hodjegan, The Simcyp population-based ADME simulator, *Expert Opin Drug Metab Toxicol* 5 (2) (2009) 211–223, https://doi.org/10.1517/17425250802691074.
[59] EMBL-EBI. (2018, 05/02/2024). *ChEMBL*. https://www.ebi.ac.uk/chembl/.
[60] Stratford, K., Nakayama, S., Jones, J., Langthaler, K., Sarfraz, M., & Jesudason, C. D. (2024). *Admet property prediction: Machine Learning: AI-Driven Drug Design*. https://www.simulations-plus.com/software/admetpredictor/.
[61] M. Jamei, D. Turner, J. Yang, S. Neuhoff, S. Polak, A. Rostami-Hodjegan, G. Tucker, Population-based mechanistic prediction of oral drug absorption, *AAPS J* 11 (2) (2009) 225–237, https://doi.org/10.1208/s12248-009-9099-y.
[62] R.S. Yang, F. Coulston, L. Golberg, Binding of hexachlorobenzene to erythrocytes: species variation, *Life Sci* 17 (4) (1975) 545–549, https://doi.org/10.1016/0024-3205(75)90088-0.
[63] Y. Lu, M. Lohitnavy, M.B. Reddy, O. Lohitnavy, A. Ashley, R.S. Yang, An updated physiologically based pharmacokinetic model for hexachlorobenzene: incorporation of pathophysiological states following partial hepatectomy and hexachlorobenzene treatment, *Toxicol Sci* 91 (1) (2006) 29–41, https://doi.org/10.1093/toxsci/kfj133.
[64] R.G. Pearce, R.W. Setzer, C.L. Strope, N.S. Sipes, J.F. Wambaugh, httk: R Package for High-Throughput Toxicokinetics, *J Stat Softw* 79 (4) (2017) 1–26. https://doi.org/10.18637/jss.v079.i04.
[65] Filaquier, C. M. (1994). *RPA 201772: Absorption, Distribution, Metabolism and Excretion in the Rat*.
[66] National Toxicology Program (NTP), TOX-77: Hexachlorobenzene (118-74-1) (Chemical Effects in Biological Systems (CEBS) Issue (2021). https://cebs.niehs.nih.gov/cebs/publication/TOX-77.
[67] D.W. Kelly, C.L. Holder, W.A. Korfmacher, T.A. Getek, J.O. Lay Jr., D.A. Casciano, J.G. Shaddock, H.M. Duhart, W. Slikker Jr., Metabolism of methapyrilene by Fischer-344 rat and B6C3F1 mouse hepatocytes, *Xenobiotica* 22 (12) (1992) 1367–1381, https://doi.org/10.3109/00498259209056688.
[68] D.W. Kelly, C.L. Holder, W.A. Korfmacher, W. Slikker Jr., Plasma elimination and urinary excretion of methapyrilene in the rat, *Drug Metab Dispos* 18 (6) (1990) 1018–1024. https://www.ncbi.nlm.nih.gov/pubmed/1981507.
[69] M.L. Gargas, M.E. Andersen, H.J. Clewell 3rd., A physiologically based simulation approach for determining metabolic constants from gas uptake data, *Toxicol Appl Pharmacol* 86 (3) (1986) 341–352, https://doi.org/10.1016/0041-008x(86)90361-3.
[70] D.J. Paustenbach, G.P. Carlson, J.E. Christian, G.S. Born, A comparative study of the pharmacokinetics of carbon tetrachloride in the rat following repeated inhalation exposures of 8 and 11.5 hr/day, *Fundam Appl Toxicol* 6 (3) (1986) 484–497, https://doi.org/10.1016/0272-0590(86)90222-8.
[71] D.J. Paustenbach, J.E. Christian, G.P. Carlson, G.S. Born, The effect of an 11.5-hr/day exposure schedule on the distribution and toxicity of inhaled carbon tetrachloride in the rat, *Fundam Appl Toxicol* 6 (3) (1986) 472–483, https://doi.org/10.1016/0272-0590(86)90221-6.
[72] J.C. Ramsey, M.E. Andersen, A physiologically based description of the inhalation pharmacokinetics of styrene in rats and humans, *Toxicol Appl Pharmacol* 73 (1) (1984) 159–175, https://doi.org/10.1016/0041-008x(84)90064-4.
[73] K.D. Thrall, M.E. Vucelick, R.A. Gies, R.C. Zangar, K.K. Weitz, T.S. Poet, D. L. Springer, D.M. Grant, J.M. Benson, Comparative metabolism of carbon tetrachloride in rats, mice, and hamsters using gas uptake and PBPK modeling, *J Toxicol Environ Health A* 60 (8) (2000) 531–548, https://doi.org/10.1080/00984100050082085.
[74] National Toxicology Program (NTP), *TR-520: 3,3',4,4',5-pentachlorobiphenyl (57465-28-8)*. (Chemical Effects in Biological Systems (CEBS), Issue (2006). https://cebs.niehs.nih.gov/cebs/publication/TR-520.
[75] National Toxicology Program (NTP), *TR-521: 2,3,7,8-Tetrachlorodibenzo-p-dioxin (1746-01-6)*. (Chemical Effects in Biological Systems (CEBS), Issue (2006). https://cebs.niehs.nih.gov/cebs/publication/TR-521.
[76] National Toxicology Program (NTP), *TR-526: TEF evaluation (Dioxin mixture) (TEFDIOXINMIX)* (Chemical Effects in Biological Systems (CEBS), Issue (2006). https://cebs.niehs.nih.gov/cebs/publication/TR-526.
[77] National Toxicology Program (NTP), *TR-530: TEF evaluation (Binary mixture; PCB 126/PCB 153) (TEFBINARYMIX)* (Chemical Effects in Biological Systems (CEBS), Issue (2006). https://cebs.niehs.nih.gov/cebs/publication/TR-530.
[78] National Toxicology Program (NTP), TR-559: PCB 118 (31508-00-6) (Chemical Effects in Biological Systems (CEBS) Issue (2010). https://cebs.niehs.nih.gov/cebs/publication/TR-559.
[79] Barash, H., R Gross E Fau - Edrei, Y., Edrei Y Fau - Ella, E., Ella E Fau - Israel, A., Israel A Fau - Cohen, I., Cohen I Fau - Corchia, N., Corchia N Fau - Ben-Moshe, T., Ben-Moshe T Fau - Pappo, O., Pappo O Fau - Pikarsky, E., Pikarsky E Fau - Goldenberg, D., Goldenberg D Fau - Shiloh, Y., Shiloh Y Fau - Galun, E., Galun E Fau - Abramovitch, R., & Abramovitch, R. (2010). Accelerated carcinogenesis following liver regeneration is associated with chronic inflammation-induced double-strand DNA breaks. *Proc Natl Acad Sci USA*, *107*, 2207-2212. https://doi.org/10.1073/pnas.0908867107.
[80] M.J. Bissell, W.C. Hines, Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression, *Nat Med* 17 (3) (2011) 320–329, https://doi.org/10.1038/nm.2328.
[81] W. Burggren, G. Martinez Bautista, T. Göpel, P. Padilla, Lack of Genetic Variation in Low Heterozygosity and Clonal Animals Creates Lack of Physiological Variation, *Physiology* 39 (S1) (2024) 1477, https://doi.org/10.1152/physiol.2024.39.S1.1477.
[82] E.L. Burrows, C.E. McOmish, A.J. Hannan, Gene–environment interactions and construct validity in preclinical models of psychiatric disorders, Prog Neuropsychopharmacol Biol Psychiatry 35 (6) (2011) 1376–1382. https://doi.org/10.1016/j.pnpbp.2010.12.011.
[83] P. Rubio Arzola, R.M. Shansky, Considering Organismal Physiology in Laboratory Studies of Rodent Behavior, Annu Rev Neurosci 45 (45) (2022) 387–402. https://doi.org/10.1146/annurev-neuro-111020-085500.
[84] B. Voelkl, N.S. Altman, A. Forsman, W. Forstmeier, J. Gurevitch, I. Jaric, N.A. Karp, M.J. Kas, H. Schielzeth, T. Van de Casteele, H. Würbel, Reproducibility of animal

research in light of biological variation, *Nat Rev Neurosci* 21 (7) (2020) 384–393, https://doi.org/10.1038/s41583-020-0313-3.

[85] M. Chassan, A. Hémonic, D. Concordet, What Matters in Piglets' Exposure to Antibiotics Administered through Drinking Water? *Antibiotics* 10 (9) (2021) 1067. https://www.mdpi.com/2079-6382/10/9/1067.

[86] A.L. Soraci, F. Amanto, M.O. Tapia, E. de la Torre, P.-L. Toutain, Exposure variability of fosfomycin administered to pigs in food or water: Impact of social rank, Res Vet Sci 96 (1) (2014) 153–159. https://doi.org/10.1016/j.rvsc.2013.12.003.

[87] E.J. Chesler, S.G. Wilson, W.R. Lariviere, S.L. Rodriguez-Zas, J.S. Mogil, Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive, Neurosci Biobehav Rev 26 (8) (2002) 907–923. https://doi.org/10.1016/S0149-7634(02)00103-3.

[88] F. Aeffner, K. Wilson, N.T. Martin, J.C. Black, C.L.L. Hendriks, B. Bolon, D. G. Rudmann, R. Gianani, S.R. Koegler, J. Krueger, G.D. Young, The Gold Standard Paradox in Digital Image Analysis: Manual Versus Automated Scoring as Ground Truth, *Arch Pathol Lab Med* 141 (9) (2017) 1267–1275, https://doi.org/10.5858/arpa.2016-0386-RA.

[89] European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC). (2003). *Derivation of Assessment Factors for Human Health Risk Assessment - Technical Report No. 86* (ISSN-0773-6347-86).

[90] J. Chen, A. Noorlander, S. Wesseling, H. Bouwmeester, N.I. Kramer, I. Rietjens, Integrating In Vitro Data and Physiologically Based Kinetic Modeling to Predict and Compare Acute Neurotoxic Doses of Saxitoxin in Rats, Mice, and Humans, *Environ Sci Technol* 57 (30) (2023) 10974–10984, https://doi.org/10.1021/acs.est.3c01987.

[91] R.A. Clewell, J.A. Leonard, C.I. Nicolas, J.L. Campbell, M. Yoon, A.Y. Efremenko, P. D. McMullen, M.E. Andersen, H.J. Clewell 3rd, K.A. Phillips, Y.M. Tan, Application of a combined aggregate exposure pathway and adverse outcome pathway (AEP-AOP) approach to inform a cumulative risk assessment: A case study with phthalates, *Toxicol in Vitro* 66 (2020), https://doi.org/10.1016/j.tiv.2020.104855.