A METHOD TO CALIBRATE CHEMICAL AGNOSTIC QUANTITATIVE ADVERSE OUTCOME PATHWAYS ON MULTIPLE CHEMICAL DOSE-RESPONSE DATA

Zheng Zhou[1], and Ullrika Sahlin[1]

2025-02-17

[1] Center for Environmental and Climate Science, Lund University

ABSTRACT

Quantitative Adverse Outcome Pathways (qAOPs) may support next-generation risk assessment by integrating New Approach Methodologies (NAMs) for derivation of relevant point of departures for hazard characterisation. To be useful, a qAOP should be chemical agnostic. We use a hierarchical statistical modeling approach to separate chemical-specific from chemical-agnostic patterns in response-response relationships between key events in an AOP. The method implements chemical-specific parameters governed by an overarching hierahical structure. Through a simulation study, we demonstrate that the hierarchical specification can capture between chemical heterogeneity, allowing for a way to evaluate if a calibrated qAOP is chemical agnostic. The model's practical utility was validated using a case study of non-genotoxic liver tumor development, successfully handling heterogeneous responses while maintaining pathway-level insights. This approach enhances qAOP generalizability while preserving chemical agnosticity, supporting reliable NAMs-based next-generation risk assessments.

1 INTRODUCTION

1.1 Background

The Adverse Outcome Pathways (AOP) framework facilitates the extrapolation of evidence from animal and non-animal methods at lower biological levels for the inference of Adverse Outcomes (AOs) at apical levels (OECD 2017; Villeneuve et al. 2024). Robust AOPs may point to a direction to Reduce, Replace and Refine (3R) animal experiments in next-generation risk assessments (Pallocca et al. 2022). The development of quantitative AOPs

(qAOPs) is one of the three stages of AOP development (OECD 2017; Villeneuve et al. 2024). The qAOP stage focuses on the development and evaluation of quantitative understandings of Key Event Relationships (KERs) (Villeneuve et al. 2024). qAOPs have been proposed to support hazard identification, derivation of points of departure (POD) and development of integrated testing and assessment approaches in next-generation risk assessments (Spinu et al. 2020; Zhou et al. 2025).

1.2 Challenge to Achieve Chemical Agnostic Adverse Outcome Pathways

Being chemical agnostic is one of the five key principle of AOPs (Villeneuve et al. 2014). This principle is defined in two ways. Firstly, the starting point of an AOP is the molecular initiating event, because the AOP framework organizes existing knowledge on the generalized patterns of toxicological and physiological mechanisms leading to the AO (Villeneuve et al. 2024). In comparison, a mode of actions (MOA) is specific to a chemical of interest and its starting point is the dose of the chemical. Secondly, an explicit assumption underlying the principle is that regardless chemical, as long as the MIE is triggered at a sufficiently severe magnitude and duration, there exists the potential to activate the downstream chain of key events (KEs) in the AOP.

However, current qAOP studies face the challenge to account for between-chemical heterogeneity. Among chemicals that have the potential to cause the same AO, there are intrinsic heterogeneity across them, such as the distinctions in chemical structures and properties, and the mechanistic pathways and KEs through which the AO is induced (Villeneuve et al. 2014). Specifically, the challenge to account for uncertainty from between-chemical heterogeneity in chemical-agnostic qAOPs are two-fold.

Firstly, qAOP models may have limited generalizability if they are based on single-chemical data. Many existing qAOP models were developed and calibrated based on the data of a specific chemical and a specific route of exposure/ assay of measurement (such as Zgheib et al. (2019);Song et al. (2023)). On the one hand, the quality of these single-chemical data could be outstanding so that the resulted qAOPs may have excellent predictive performance. On the other hand, however, the extent to which the model would be valid for the inference of other chemicals that can cause the same AO was not necessarily considered in model development. Consequently, these single-chemical qAOPs may provide limited implication for the underlying AOP. This problem of limited generalizability greatly restricts the applicability of these qAOP models for the prospective prediction of new chemicals for regulatory and industrial purposes.

Secondly, there is a lack of research to account for between-chemical heterogeneity in qAOP models on multi-chemical data. Although some qAOPs studies have incorporated multi-chemical data (such as (Burgoon et al. 2020; Furxhi et al. 2019; Gou et al. 2023)), their approaches may have limited capacities to isolate the chemical-agnostic responses associated with the AOP from the chemical-specific responses pertaining to individual chemical heterogeneity. Multi-chemical data may include a varying level of between-chemical heterogeneity in terms of chemical structure, chemical and physical properties, and most importantly activities of KEs from the AOP of interest. When heterogeneity is high, it is possible that a dose that is sufficiently severe for one chemical to induce the full chains of KEs and the AO, is insufficient to trigger the MIE in another chemical. Statistically, a model could be biased by the extreme data points if heterogeneity is not properly accounted (McElreath 2018). Various approaches have been employed for qAOP modeling

with multi-chemical data, including machine learning, Bayesian network and parametric response-response models (Burgoon et al. 2020; Furxhi et al. 2019; Gou et al. 2023). In the parametric models employed by these studies, the same set of parameters, such as regression coefficients, were estimated for all chemicals with no consideration to heterogeneity. Furthermore, there was limited discussion on the extent to which the results of these studies could be biased by high heterogeneity.

Therefore, it is useful to develop an approach to measure the heterogeneity level for the calibration of chemical-agnostic qAOPs. Hierarchical structure is an established approach in applied statistics to tackle heterogeneity in parametric models (Gelman et al. 2013; McElreath 2018). There have been successful applications of hierarchical structures in dose-response analysis (Allen et al. 2020; Rota et al. 2010). Yet there is lack of exploration on how to implement hierarchical structures in qAOP models, specifically parametric response-response models.

1.3 Study Objectives

To address the limitations in current qAOP studies, the objective of this study is to develop a statistical approach for qAOP models to identify and calibrate chemical-agnostic properties. To achieve this objective, we propose a qAOP with hierarchical structures to separate the response-response patterns of chemical-specific heterogeneity from that of chemical-agnostic pathways. The difference in predictive performances between models with and without hierarchical structures is proposed to be a *post hoc* measure of heterogeneity level in a chemical-agnostic pathway.

2 METHODS

2.1 Quantitative Adverse Outcome Pathway Model

Our qAOP has a dose-response (part 1) and a response-response (part 2) as in Equation (1). Subscripts 1 and 2 correspond to KE_{up} and KE_{down} respectively. In part 1, the outcome of KE_{up} is modeled as a random variable Y_1. It follows a distribution F_1, whose parameter values are computed from a dose-dependent function $f_1(d)$. Part 2 specifies that the outcome of KE_{down} is a random variable Y_2 following a distribution F_2, where the parameters are computed from a function $f_2(Y_1)$.

$$\begin{aligned} Y_1 &\sim F_1\big(f_1(d)\big) \\ Y_2 &\sim F_2\big(f_2(Y_1)\big) \\ \Rightarrow Y_2 &\sim F_2\left(f_2\left(F_1\big(f_1(d)\big)\right)\right) \end{aligned} \quad (1)$$

We employed the benchmark dose (BMD) methodology for the modeling of Y_1 and Y_2. For a comprehensive description of the BMD methodology, we recommend the User Guide (EPA 2022b) and Technical Guidance (EPA 2022a) of U.S. EPA Benchmark Dose Software (BMDS). Outcomes could be continuous or categorical. For the purpose of this study, it was assumed that Y_1 and Y_2 are continuous and binary, respectively. For the parameterization of F, the following assumption from BMDS (EPA 2022a) were followed as in Equation (2): 1) The outcome of KE_{up} follows a normal distribution at any doses d, where the mean $f(d)$ is dose-dependent with parameters θ_1; 2) The outcome of KE_{down} follows a Bernoulli distribution at any levels of Y_1, where the probability is dependent on Y_1 with parameters θ_2.

$$\begin{aligned} Y_1 &\sim N(f_1(d, \theta_1), \sigma_1) \\ Y_2 &\sim Bern\big(f_2(Y_1, \theta_2)\big) \end{aligned} \quad (2)$$

We chose a continuous-Hill function from BMDS (EPA 2022a) (Equation (3)) for f_1 of the dose-response part. In this function, a is control response; b is maximum change, $b > 0$ for increasing trend, $b < 0$ for decreasing trend; $c > 0$ is the dose at which half-maximal change occurs; $g > 0$ is sigmoid power. Hill function is flexible and biologically relevant (EPA 2022a; Weiss 1997). A dichotomous-Hill function from BMDS (Equation (3))(EPA 2022a) was adapted for f_2 of the response-response part, where $0 < v <= 1$ is maximum probability; $0 <= b < 1$ is background risk so that vb is background probability; a is intercept; $g >= 0$ is power. The flexibility of Hill functions allows the modeling of a potential threshold in the dose-response and response-response relationship. The impact of our choices on functional forms are discussed in section 4 Discussion.

$$\begin{aligned} f_1(d) &= a + b * \frac{d^g}{c^g + d^g} \\ f_2(Y_1) &= v \frac{1 + b * e^{-(a + g\,log(Y_1))}}{1 + e^{-(a + g\,log(Y_1))}} \end{aligned} \quad (3)$$

2.2 Hierarchical Structure

The highlight of this study is the integration of hierarchical structures with qAOP models to separate between chemical-agnostic and chemical-specific patterns in the data of multiple chemicals.

To demonstrate the specification with and without hierarchical structures, consider the dichotomous-Hill function for part 2 (Equation (3)). Consider that a multi-chemical data is clustered by chemical index $k = 1 \ldots K$. The specification without hierarchical structures is

thereby referred as the "*flat*" specification. It assumes the same set of parameters $\{v, b, a, g\}$ across K chemicals. In contrast, the function could be integrated with hierarchical structures on parameters to properly reflect our assumption of clustering in data (Equation (4)), simply referred as "*hierarchical*".

$$\begin{gathered} Y_{2_{k,i}}\ Bern\left(f_2\left(Y_{1_{k,i}}\right)\right) \\ f_2\left(Y_{1_{k,i}}\right) = v_k \frac{1 + b_k * e^{-\left(a + g\,log\left(Y_{1_{k,i}}\right)\right)}}{1 + e^{-\left(a + g\,log\left(Y_{1_{k,i}}\right)\right)}} \quad (4) \\ v_k \sim beta(\alpha_v, \beta_v) \\ b_k \sim beta(\alpha_b, \beta_b) \end{gathered}$$

where $i = 1 \ldots N_k$ is index of observations per chemical. The hierarchical specification includes a within-cluster component and a cross-cluster component. The within-cluster component corresponds to the effects specific to the heterogeneity of each chemical. Therefore, the within-cluster parameters include random effects that are the same within each chemical but varies across chemicals, represented by chemical-specific values for parameters v_k & b_k. The cross-cluster component corresponds to the effects from chemical-agnostic pathways. They are the same across chemical, including parameters a & g and hyperparameters $\alpha_v, \beta_v, \alpha_b, \beta_b$. The chemical-specific values of v_k & b_k are governed by the corresponding hyperparameters, representing the influence from the pathway that is chemical agnostic.

2.3 Data Requirements

Our proposed qAOP model implies two key data requirements: 1) it consists of two dose-response data sets for the two responses respectively and 2) there are $K > 0$ chemicals

shared between the two sets. Two dose-response sets are needed pertaining to the impact of the dose on the responses of up- and down-stream KEs.

High-quality data is crucial for model development and validation. The toxicological data of the highest quality is from randomized controlled trials, which provide unbiased estimates of the impact of doses on the response (Zhou et al. 2025). Data from observational studies, i.e., without randomized controlled trials, could be considered if experimental data is not available. However, potential risks of bias in observational data lead to low weights of evidence (Meek et al. 2014; Zhou et al. 2025), which may impair the robustness of the qAOP. To enable the consideration of between-chemical heterogeneity, both data sets must include multiple, shared chemicals.

With the satisfaction of the key requirements, the data quality could be further enhanced if the following criteria are met. If the two sets came from the same experiment, they may have the same applicability domain (such as an *in vivo* route of exposure or an *in vitro* assay), and dosage setups (i.e. dose groups). If the two data sets were collected by separate experiments or from different studies, the shared chemicals must have the same, or at least, comparable applicability domain. In this case, it is not necessary for the dosage setups to completely match between the two sets. Doses must be converted into the same unit, such as an average daily dose (equation in the Supplementary Information). When data on individual subject's response is not available, summary information of the response could be used as approximations. Therefore, if individual data is available, it is possible but not recommended to aggregate them in the summarized format.

2.4 Model Implementation

Parameter estimation and model validation is performed under the Bayesian framework. Posterior distributions of parameters are estimated using a Markov Chain Monte Carlo (MCMC) sampler through the R interface of stan (Team 2022), which adopts the no-U-turn sampler of the Hamiltonian Monte Carlo algorithm. The target log-density functions and full scripts for stan are provided in the SI. Prior specifications were evaluated through prior predictive distribution (Wesner and Pomeranz 2021). For the MCMC sampler, we included four chains, each consisting of 5000 warm-up iterations and 5000 sampling iterations. The performance of MCMC samplers were evaluate for chain convergence and effective sample size, using the $\hat{R}$ and effective sample size statistics respectively.

2.5 Heterogeneity Measures

The heterogeneity level of a multiple-chemical data is represented by Cohen's f (Cohen 2013) in Equation (5). Cohen's f becomes the ratio of between-chemical variance to within-chemical variance.

$$ES = \sqrt{\frac{SSB}{SST - SSB}}$$
$$SSB = \sum_{j=1}^{K} n_j \left(\mu_j - \hat{\mu}\right)^2 \quad (5)$$
$$SST = \sum_{i=1}^{N} (y_i - \hat{\mu})^2$$

where SSB and SST are between-chemical and total squared variance, respectively; $\hat{\mu}$ is global mean of the data; j is chemical index, n_j and μ_j are the size and mean per chemical.

The value of Cohen's f can be interpreted as follows (Cohen 2013): when f=0, there is no between-chemical differences; when f>0, 0.1 = low differences, 0.25 = medium differences, 0.4 = big differences and 0.7 = large differences. In this study, the level of Cohen's f is interpreted as the level of heterogeneity.

In addition to Cohen's f, heterogeneity level is also identified with a *post hoc* metric by comparing the predictive performances between the flat and hierarchical specifications. The performance of each specification is evaluated by the expected log predictive density (elpd, Equation (6)), because it allows the comparison of posterior predictive performances between non-nested, non-linear models (Vehtari et al. 2017; Vehtari and Ojanen 2012).

$$elpd = \sum_{i=1}^{n} \int p_t(\tilde{y}_i) logp(\tilde{y}_i|y)\, d\tilde{y}_i \quad (6)$$

where $p_t(\tilde{y}_i)$ is the distribution for the true data-generating process for $\tilde{y}_i$. In this study, we employed the Pareto Smoothed Leave-one-out Cross Validation (PSIS-LOO) method to approximate $p_t(\tilde{y}_i)$. This approximation was implemented using the '*loo*' package in R (Vehtari et al. 2024). Within this paper, $elpd_{PSIS-LOO}$ is simply referred as elpd.

Consequently, the differences in elpd between the hierarchical and flat specifications is the *post hoc* measure of heterogeneity that $\Delta elpd = elpd_{hier} - elpd_{flat}$. According to the definition of elpd (Vehtari 2024; Vehtari et al. 2017), positive differences show that the hierarchical specification performs better than the flat specification in terms of elpd. This would suggest that the between-chemical heterogeneity in data is large enough for the need of hierarchical structures. The difference in model performance was also evaluated using Akaike Information Criteria (AIC, Akaike (2011)) in comparison to $elpd_{PSIS-LOO}$,

which is considered as the conventional choice for model comparison (Wasserman 2000). A negative AIC difference indicates that the hierarchical specification performs better than the flat one in terms of AIC.

2.6 Simulation Study

A simulation study was performed to evaluate our qAOP. The procedure of the simulation study is demonstrated in Figure 1. Full scripts for simulation and analyses are available in the SI.

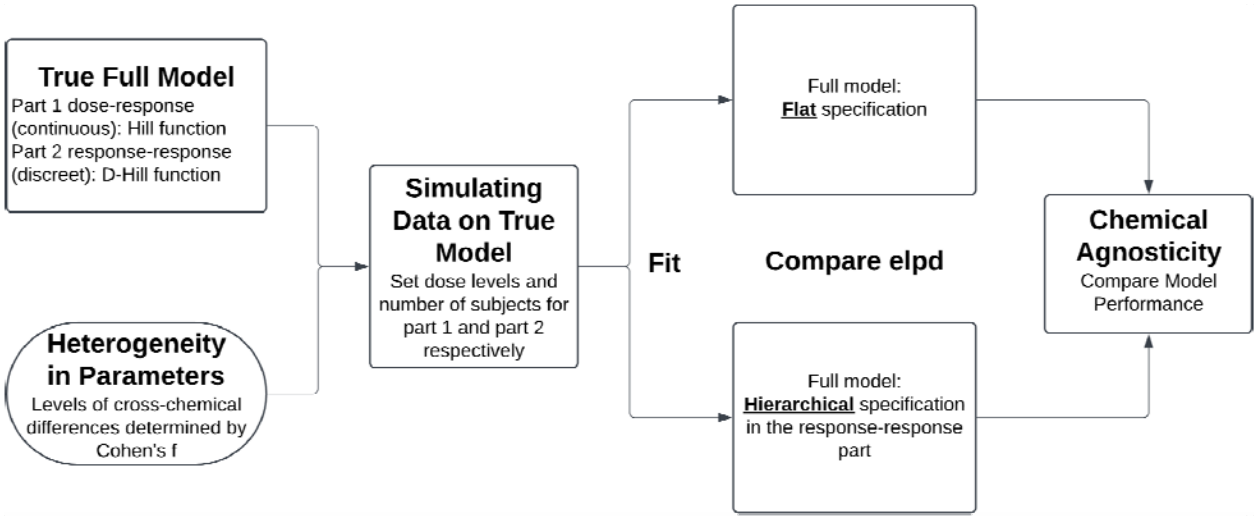

Figure 1: Diagram of Simulation Study

Data was simulated to include ten chemicals. Four dose levels were randomly sampled per chemical within the range of 0 to 1 for and as and , respectively. Each dose group included of ten subjects. Functions and from (3) were used as the true model for simulation.

To better evaluate uncertainty in the response-response part, fixed dose-response parameters were applied across chemicals. Moreover, fixed values were applied for

response-response parameters $a\&g$ so that the curve incorporates a threshold-like shape. Four levels of between-chemcial heterogeneity was specified in the response-response parameters as none, low, medium and high. The response-response parameters were set so that the Cohen's f in the simulated data match that: None <.001, Low = 0.1, Medium = 0.2, High >= 0.4. Our simulation includes 100 iterations at each heterogeneity level to account for random noises.

The simulated dose-response data included dose d_1, KE_{up} outcomes at d_1 as $Y_1|d_1$ and chemical indexes. The simulated response-response data consisted of dose d_2, KE_{up} and $KEdown$ outcomes at d_2 as $Y_1|d_2$ and $Y_2|d_2$ respectively, and chemical indexes. The simulated data were fitted by flat and hierarchical specifications separately. Since the functional forms of the true models were known, model selection or averaging was omitted from the simulation study. The difference in the performance between the flat and hierarchical specifications was evaluated by elpd and AIC values.

2.7 Validation in by Real Data

The comparison models with and without a hierarchical structure was further validated using real non-animal data.

3 RESULTS

3.1 Simulation Study

The simulated dose-response and response-response data exhibit the desired patterns specified in the true model. The dose-response curves are almost identical across chemicals

in terms of shape and positions (Fig S1 in the SI) with random noises, which shows that uncertainty in the dose-response part was properly controlled. The shape of response-response curves of simulated data depends on heterogeneity levels (Fig S2-S5). When heterogeneity level = None, the response-response curves of different chemicals have identical shapes, positions and values of threshold (Fig S2); In contrast, when heterogeneity level = High, the curves exhibit distinctive shapes and positions (Fig S5). The location of threshold values, however, were set to be consistent.

Both of our flat and hierarchical models provided robust fittings to the simulated data at all heterogeneity levels. With 10,000 steps in the MCMC, the chains were well converged with a good number of effective sample sizes. More specifically, at heterogeneity level = None, the MCMC has good convergence and is free of overfitting problems. When heterogeneity level is high, the fittings by both specifications are acceptable (Fig S5 & S6).

Table 1: Difference in Predictive Performances at Four Heterogeneity Levels

Hetero	Cohen_f	elpd_diff	AIC_diff
0=None	<.001	-0.35(2.39)	22.88
1=Low	0.1	2.1(1.8)	20.74
2=Medium	0.22	13.82(3.84)	3.49
3=High	0.7	28.11(6.36)	-11.53

Cohen's f: ratio of between-chemical variance to within-chemical variance; elpd_diff: standard error in parenthesis, positive value favors hierarchical models; AIC_diff: negative value favors hierarchical models

The difference in predictive performances between the flat and hierarchical specifications is summarized (Table 1). These results provide two implications.

Firstly, as heterogeneity level increases, both elpd and AIC differences support that the hierarchical specification becomes more advantageous. At the level 0=None (Cohen's f<.001), the flat specification is better than the hierarchical one. The reason is that both elpd and AIC penalize for the number of parameters. With no between-chemical heterogeneity, one same of parameter is sufficient for multiple chemical, which is more consistent with the flat specification. Despite free of overfitting, the hierarchical specification is penalized heavier. Consequently, the flat model is better when between-chemical heterogeneity is absent At the level 3=High, elpd and AIC differences unanimously support the hierarchical specification can capture between-chemical heterogeneity much better than the flat one, which is also support by the fitted response-response curves (Fig S5&S6).

Secondly, the *post hoc* method of model evaluation matters. Both elpd and AIC differences support the trend described above. However, they are not in complete agreement across heterogeneity levels. Elpd difference indicates that the hierarchical specification starts to be superior than the flat one from the heterogeneity level 1=Low. In contrast, AIC difference indicates that the hierarchical specification is stronger only at the level 3=High. This is because the two methods employ different metrics to evaluate predictive performances. Elpd utilizes pointwise predictive densities (Equation (6)), while AIC is a point estimate aggregating across observations. Furthermore, the degree of penality on overfitting is different between elpd and AIC.

3.2 Validation by Real Data

4 DISCUSSIONS

The difference between elpd and AIC results supports our decision to introduce AIC as a comparison. AIC has been a conventional approach for frequentist model evaluation. It may not be optimal for the evaluation of Bayesian models (Vehtari and Ojanen 2012). In comparison, $elpd_{PSIS-LOO}$ better evaluates models under Bayesian framework (Vehtari et al. 2017). Therefore, we suggest that elpd is superior than AIC as a *post hoc* indicator of heterogeneity level in the data.

5 CONCLUSIONS

Allow a way to evaluate if a calibrated qAOP is chemical agnostic or not.

This consistent pattern across both metrics validates our approach of using hierarchical structures to handle high heterogeneity in multi-chemical data.

REFERENCES

Akaike H. 2011. Akaike's information criterion. International encyclopedia of statistical science 25–25.

Allen B, Shao K, Hobbie K, Mendez Jr W, Lee JS, Cote I, et al. 2020. Bayesian hierarchical dose-response meta-analysis of epidemiological studies: Modeling and target population prediction methods. Environment International 145; doi:10.1016/j.envint.2020.106111.

Burgoon LD, Angrish M, Garcia-Reyero N, Pollesch N, Zupanic A, Perkins E. 2020. Predicting the probability that a chemical causes steatosis using adverse outcome pathway bayesian networks (AOPBNs). Risk Analysis 40:512–523; doi:10.1111/risa.13423.

Cohen J. 2013. *Statistical power analysis for the behavioral sciences*. routledge.

EPA US. 2022a. *Benchmark dose technical guidance*.

EPA US. 2022b. *BMDS version 3.3 user guide (oct 2022)*. U.S. Environmental Protection Agency:Washington, DC.

Furxhi I, Murphy F, Poland CA, Sheehan B, Mullins M, Mantecca P. 2019. Application of bayesian networks in determining nanoparticle-induced cellular outcomes using transcriptomics. Nanotoxicology 13:827–848; doi:10.1080/17435390.2019.1595206.

Gelman A, Stern HS, Carlin JB, Dunson DB, Vehtari A, Rubin DB. 2013. *Bayesian data analysis*. Chapman; Hall/CRC.

Gou X, Ma C, Ji H, Yan L, Wang P, Wang Z, et al. 2023. Prediction of zebrafish embryonic developmental toxicity by integrating omics with adverse outcome pathway. Journal of Hazardous Materials 448:130958; doi:10.1016/j.jhazmat.2023.130958.

McElreath R. 2018. *Statistical rethinking: A bayesian course with examples in r and stan*. Chapman; Hall/CRC.

Meek M, Palermo CM, Bachman AN, North CM, Jeffrey Lewis R. 2014. Mode of action human relevance (species concordance) framework: Evolution of the bradford hill considerations

and comparative analysis of weight of evidence. Journal of Applied Toxicology 34:595–606; doi:10.1002/jat.2984.

OECD. 2017. Guidance document for the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA).

Pallocca G, Moné MJ, Kamp H, Luijten M, Van de Water B, Leist M. 2022. Next-generation risk assessment of chemicals - rolling out a human-centric testing strategy to drive 3R implementation: The RISK-HUNT3R project perspective. Alternatives to Animal Experimentation: ALTEX 39:419–426; doi:10.14573/altex.2204051.

Rota M, Bellocco R, Scotti L, Tramacere I, Jenab M, Corrao G, et al. 2010. Random-effects meta-regression models for studying nonlinear dose-response relationship, with an application to alcohol and esophageal squamous cell carcinoma. Statistics in medicine 29:2679–2687; doi:10.1002/sim.4041.

Song Y, Zheng K, Brede DA, Gomes T, Xie L, Kassaye Y, et al. 2023. Multiomics point of departure (moPOD) modeling supports an adverse outcome pathway network for ionizing radiation. Environmental Science & Technology 57:3198–3205; doi:10.1021/acs.est.2c04917.

Spinu N, Cronin MTD, Enoch SJ, Madden JC, Worth AP. 2020. Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. Archives of Toxicology 94:1497–1510; doi:10.1007/s00204-020-02774-7.

Team SD. 2022. RStan: The r interface to stan.

Vehtari A. 2024. Cross-validation FAQ.

Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P-C, Paananen T, et al. 2024. Loo: Efficient leave-one-out cross-validation and WAIC for bayesian models.

Vehtari A, Gelman A, Gabry J. 2017. Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27:1413–1432; doi:10.1007/s11222-016-9696-4.

Vehtari A, Ojanen J. 2012. A survey of bayesian predictive methods for model assessment, selection and comparison. Statistics Surveys 6:142–228; doi:10.1214/12-SS102.

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, et al. 2014. Adverse outcome pathway (AOP) development i: Strategies and principles. Toxicological Sciences 142:312–320; doi:10.1093/toxsci/kfu199.

Villeneuve D, Meek B, Viviani B, Burgdorf T, LaLone C, O'Brien J, et al. 2024. AOP developers' handbook 2.7.

Wasserman L. 2000. Bayesian model selection and model averaging. Journal of Mathematical Psychology 44:92–107; doi:10.1006/jmps.1999.1278.

Weiss JN. 1997. The hill equation revisited: Uses and misuses. The FASEB Journal 11:835–841; doi:10.1096/fasebj.11.11.9285481.

Wesner JS, Pomeranz JPF. 2021. Choosing priors in bayesian ecological models by simulating from the prior predictive distribution. Ecosphere 12; doi:10.1002/ecs2.3739.

Zgheib E, Gao W, Limonciel A, Aladjov H, Yang H, Tebby C, et al. 2019. Application of three approaches for quantitative AOP development to renal toxicity. Computational Toxicology 11:1–13; doi:10.1016/j.comtox.2019.02.001.

Zhou Z, Pennings JLA, Sahlin U. 2025. Causal, predictive or observational? Different understandings of key event relationships for adverse outcome pathways and their implications on practice. Environmental Toxicology and Pharmacology 113:104597; doi:10.1016/j.etap.2024.104597.