
Article:
Gosling, JP (2019) The importance of mathematical modelling in chemical risk assessment and the associated quantification of uncertainty. Computational Toxicology, 10. pp. 44-50. ISSN: 2468-1113

https://doi.org/10.1016/j.comtox.2018.12.004

The importance of mathematical modelling in chemical risk assessment and the associated quantification of uncertainty

John Paul Gosling[a,]

[a]*School of Mathematics, University of Leeds, Leeds, LS2 9JT, United Kingdom*

Abstract

Computational models pervade modern toxicology and are becoming an accepted part of chemical risk assessments. Mathematical and statistical tools are versatile enough to capture information from wide arrays of existing data and from our mechanistic understanding of human biology and chemical reactions. They are more accessible than ever given the number of readily available guidance documents and software packages. In the present article, we will highlight the utility of modelling for next generation risk assessments whilst emphasising the importance of characterising and reporting uncertainty. The concepts herein are the foundations for a paradigm shift in toxicology where transparency about scientific understanding replaces faith in animal models.

Keywords: *in silico* predictions, mathematical modelling, next generation risk assessment, uncertainty

1. Introduction

Mathematical and statistical models are becoming popular tools for toxicologists when performing chemical risk assessments. However, there is still a long way to go for these tools to form the core of such assessments due to the general reluctance to accept new approaches [1, 2]. Some of this reluctance to acceptance is due to the fact that mathematical and statistical models cannot fully replicate the complexity of human biology and its interactions with chemicals. Analogously, mice, rats and other animals are not the same as humans, and, where we are able to measure effects in humans, the experimental cohort will almost always be a tiny subset of the population of interest and not fully representative of the people that we would like to protect. Realising the imperfections of our model systems, be them animal or mathematical, is an important part of making them fit-for-purpose as part of a chemical risk assessment. Despite the imperfections, mathematical models should still be attractive to risk assessors

Email address: j.p.gosling@leeds.ac.uk (John Paul Gosling)

because the costs of running mathematical models as computer-based experiments or statistical prediction models are far less than the costs of laboratory experimentation.

In a laboratory, a researcher would hope that their experiments can be used to test plausible hypotheses that accord with their scientific understanding. Similarly, mathematical models provide a window on what the modeller believes is occurring in a system of interest. In our context, we could consider the process of investigating a mathematical model as being the same as when experimenting in the laboratory. When an experiment is performed, the experimenter believes that the observed response (or lack of it) is informative about the human response. However, this is a subjective judgement, and risk assessors will have their own views on which experiments are most informative for different toxicological endpoints. A more detailed exposition of what is covered by the phrase "mathematical model" is given in Section 2.

To aid modern chemical risk assessment, we must appreciate the uncertainty in the transition from experimental and computational results to the risk assessment endpoints of interest. Our motivation for addressing the uncertainty in using mathematical models comes from the fact that understanding uncertainty is fundamental for defensible risk management. In this, there is a distinction to be made between aleatory and epistemic uncertainty: aleatory uncertainty covers the irreducible randomness that we are used to seeing when we sample from populations or when we repeat experiments and epistemic uncertainty covers the lack of knowledge that can be reduced by collecting further evidence. Chemical risk assessment has both types of uncertainty: for instance, there are the experimental variabilities, which give rise to aleatory uncertainty, and unknowns regarding the relevance of experimental results to populations we are trying to predict, which give us epistemic uncertainty. Many other fields have embraced uncertainty quantification as a crucial part of their risk assessments: for example, nuclear waste containment [3], flood defences [4] and exotic diseases in livestock [5]. Information on the probabilities of alternative outcomes that stem from quantitative uncertainty analyses is necessary to enable decision makers to choose courses of action that increase the chances of favourable outcomes, which, in our setting, protect against toxic effects in the population of interest. Suggestions of methods relevant to toxicology for characterising and dealing with uncertainty are given in Section 3. In the present article, we focus on chemical risk assessment for human populations, but the methods and principles stated could easily transfer to ecotoxicological or other chemical risk settings.

In Section 4, we will offer a framework for assimilating mathematical models into chemical risk assessments whilst pointing at attempts to fulfill part of this aim and referencing relatively new guidelines and methodologies that will aid risk assessors in implementing the framework. The suggestions are with reference to next generation risk assessments (NGRAs) [6] and building mathematical models that are consistent with adverse outcome pathway (AOP) approaches to understanding toxicities at the population level [7]. Also, in Section 4, we link these concepts with mathematical model building and uncertainty quan-

tification where we have lots of data that are not necessarily directly about the chemical or human endpoint of interest.

2. Mathematical modelling in toxicology

Before discussing how models and better accounting for uncertainty can aid chemical risk assessment, we will discuss the different types of mathematical model available to risk assessors. One reason for spelling out the different model types here is because terms such as "computer-based models" [8], "network models" [9], "prediction models" [10] and "*in silico* models" [11, 12] have been used in many settings, and, in some cases, interchangeably for very different types of models that are used for making toxicity predictions. In Table 1, we have provided examples of the terms used to describe models that can be broadly categorised as one of three model types: data-driven, statistical and mechanistic. Also, in Table 1, some of the key features of these model types are given alongside examples of the models being employed in the toxicology literature. The division into three distinct categories here is forced to help highlight the different model types, and there is a continuum of possibilities as the mathematical methods go from being mechanistic to data driven. The move from data-driven to statistical to mechanistic models sees more structure imposed on the model that is thought to mirror the biological and chemical realities of the system being modelled, and there is also a greater burden on the user to make more specifications about the system. For instance, a machine learning approach might require access to a database of toxicities whereas a mechanistic approach would need the specification in mathematical terms of how such toxicities occur.

Within all the groups of models in Table 1, we have included quantitative structure-activity relationship models (QSARs), which are classification or prediction models that are built using databases that include chemical properties and toxicological outcomes. The reason they have been included in all is that they are not always exclusively built using relationships discovered in the data: information about relationships between chemical properties and toxicological end-points that is external to the data can also be included and can be the driving force in predictions. The interpretation accords with the OECD principles for the validation of QSARs [23]. The OECD principles require information on the defined endpoint for a regulatory purpose, the algorithm, the applicability domain of the model, the predictive ability of the model, and a mechanistic interpretation. These same expectations should be carried forward for more general mathematical models in chemical risk assessment, and this will be discussed in Section 4.

A mechanistic model is traditionally much more prescribed than its data-driven or statistical counterparts in that it will be based on some appreciation of a system's behavior that can be expressed in some mathematical equations that may stem from physical laws (conservation of mass, for example). When developing mechanistic mathematical models, it is important that potential users understand the behaviour of the models and, in particular, the influence that

Table 1: A non-exhaustive list of *in silico* models and some of their key features.

Type of model	Some descriptors that can be associated with this model type	Key features	Relevant examples
Data-driven	Machine learning, data mining, deep learning, neural network, QSAR, support vector machines, random forest.	Relationships are derived through discovery of patterns in the combined predictor and endpoint data set. Large number of predictor variables used to improve performance. Able to handle large amounts of data and various types of data.	[13], [14], [15].
Statistical	Regression, correlation-based, classification, read-across, QSAR, structural-alert-based, dose-response, decision trees, Bayesian belief networks.	Relationships are hypothesised between various predictor variables and the endpoints of interest. Predictor variables can be selected if importance is known *a priori*. Performance is usually determined through sensitivity and specificity for classification or goodness-of-fit for regression.	[16], [17], [18], [19].
Mechanistic	Process-based, biology-based, structural-alert-based, read across, network, agent-based, QSAR, physiologically-based kinetic, Petri-net, mathematical.	Relationships between predictor variables and the endpoints of interest are governed by mathematical equations or sets of rules. Known relationships are encoded in the implemented model. Performance is usually determined through goodness-of-fit or agreement with existing data sources.	[20], [21], [22].

different inputs have on the models' outputs. This can be made easier if the model building is guided by an adverse outcome pathway (AOP) [10] alongside biokinetic considerations [24]. An AOP gives the steps from a molecular initiating event to a population-level effect and can act as a blueprint for building a mechanistic model whilst highlighting the gaps in the evidence and the importance of the model inputs. Indeed, the use of molecular initiating events to motivate choices in *in silico* modelling is common place [25, 26]. By understanding the relative influence of the inputs, the model builders and subject-area experts are able to focus their research efforts on improving parts of the model that have a significant impact on the outputs that matter. Apart from having a formal mechanistic understanding of the input-output relationship, model builders can also employ statistical techniques to gauge the relative importance of each input. This type of exploration is traditionally done through input screening [27, 28] or sensitivity analysis techniques [29, 30]. As part of this process, it is important to verify that the model is encoded in the way that was meant from the underlying mathematical equations and relationships. It can also be beneficial to identify the inputs to that model that have an impact on the model's outputs. If we can identify such important inputs, we can determine which inputs need to be determined more carefully and which inputs we have the best chance of learning about in a parameter estimation scheme. In the context of physiologically-based kinetic (PBK) modelling, [31] provides a workflow for conducting sensitivity analyses; [32] shows the utility of investigating the sensitivity of risk assessment outcomes to modelling choices for a model of exposure to food allergens; a more general discussion that is relevant for mechanistic biology models is given in [33].

Typically, chemical risk assessors will not be experts in operating mathematical models or processing the models' results. By providing links to mechanistic understanding, as suggested in the OECD principles, or explicitly linking the model to an AOP, the assessor will be able to see the scientific basis of the model predictions and this will aid confidence. However, understanding the model's single prediction is not enough: the assessor must also be able to understand the lack of certainty in the predictions.

3. Uncertainty in chemical risk assessment

Mathematician and pioneer of probability theory Jacob Bernoulli wrote that "it is utterly implausible that a mathematical formula should make the future known to us, and those who think it can would once have believed in witchcraft" [34]. This remains true today in our context: it is unrealistic to believe that a mathematical or statistical model can completely capture the complexity of toxic effects and be able to predict population-level effects with absolute accuracy. As such, we must consider the gap between our models' predictions and reality because we need to inform decisions that occur in the real world rather than the modelled world. The Codex Alimentarius Commission, which provides an international forum for food safety issues, published a set of working principles for risk analysis that includes the following pertinent statement [35]:

> "Constraints, uncertainties and assumptions having an impact on the risk assessment should be explicitly considered at each step in the risk assessment and documented in a transparent manner. Expression of uncertainty or variability in risk estimates may be qualitative or quantitative, but should be quantified to the extent that is scientifically achievable."

The characterisation and quantification of uncertainty is important to give risk managers and decision makers confidence in their actions. The typical provision of confidence intervals or standard errors is not enough because these constructs inform us about experimental or population variability rather than uncertainty in the end-point of interest. Statistics has played a major role in the replacement, refinement and reduction of animal testing in scientific research. Classical experimental design is ubiquitous, and often mandatory, in the set-up of experiments involving animals [36, 37]. Regression modelling, and more recently machine learning techniques, are used to draw inferences from and make predictions based upon results from such experiments [38, 39]. Non-probabilistic methods such as those stemming from Dempster-Shafer theory have also been used in an attempt to capture uncertainties around QSAR modelling [40]. Using statistical methods to capture the multitude of uncertainties that can occur in risk analyses has been recommended in a number of influential guidance documents that are relevant for chemical risk assessment. One of the most recent and most comprehensive is [41] where it is stated that "uncertainty analysis is an integral part of scientific assessment" and very many methods for capturing and characterising uncertainty are given with some practical examples.

Returning to the types of models that are being used in chemical risk assessment, the model builder believes that the mathematically-modelled effect is informative about the human response. Of course, this is a subjective judgement. Also, like when applying an experimental protocol, a modeller can catalogue the choices that are made when building the model and this allows others the opportunity to challenge and improve the model in the future. The assumptions and compromises that are made give rise to various forms of uncertainty in the gap between the model and reality. Uncertainty appears in many guises including measurement error, lack of knowledge about the input parameters, discrepancy between the models and reality, and uncertainty about the linkage between the models and *in vitro* data sets. The preceding list is not exhaustive and a more detailed listing is given in [41] and specifics for PBK models are given in [42] with guidance for PBK models currently being developed by the OECD. For data-driven and statistical models, there may be other uncertainties stemming from the coverage of the training sets and the extent of their applicability domains. Of course, sensitivities and specificities of classification approaches can be evaluated, but, again, subjective judgments will be needed to characterise the uncertainty in extrapolations.

Despite the guidance and apparent importance of uncertainty characterisation, uncertainty analysis is often thought of as an addition to the end of a study: an experiment is performed or a mathematical model is run, and, then,

we consider the uncertainty in the results. Far more value can be gained for both the investigator and the ultimate risk manager if uncertainty is accommodated from the start of the process. Of course, it is unrealistic to believe that all uncertainties can be quantified, but, if we are transparent in our analyses, we can highlight areas where scientific understanding is lacking, where further research is warranted or where a risk manager will need to make a judgement about acceptable levels of risk. Further guidance on when mathematical models will be fit-for-purpose and the types of uncertainty we may encounter (including unquantifiable uncertainties) is given in [43].

The uncertainty table approach, as described in [44] and applied in [45], gives a mechanism for capturing uncertainties from the start of the process. An uncertainty table is a tool that provides transparency about the uncertainties in a risk assessment where the uncertainties and assumptions are listed and their impacts are qualitatively judged. This method is not a replacement for more rigorous methods for uncertainty modelling: it is a tool to help lay bare the assumptions in the process and to assess the impact that removing the assumptions would have. An uncertainty table in its simplest form consists of just two columns: the first is populated by the list of uncertainties and the second is used to capture judgements about the potential effects of formally accounting for the uncertainty in a probabilistic risk assessment. Table 2 shows the layout of an uncertainty table with an extra column added to capture the evidence that supports the use of the assumption and information about its impact if it were removed. The column headed direction and magnitude gives the modeller an opportunity to qualitatively assess the impact of the assumptions on the risk end-point. Further extensions and discussions of how to complete uncertainty tables are given in [44] and [46].

In the context of mathematical modelling to aid chemical risk assessment, model uncertainty can be assessed by evaluating the appropriateness of the underlying assumptions. For each assumption, the evidence to support or refute it is collected and the strength of the evidence would also be qualitatively evaluated. In addition, the impact of the assumption on the discrepancy between model prediction and reality as well as the uncertainty regarding this impact are also judged by the experts. The simplicity of the uncertainty table approach means that it can be an effective tool for keeping track of the uncertainties and assumption whilst guiding research directions and more formal quantification of uncertainty during the risk assessment. In addition to this, when the risk assessment is completed, the unresolved uncertainties (that is, the uncertainties that have not been removed or formally quantified) can be laid out for users of the risk assessment in a transparent fashion.

In most cases, it will not be sufficient to give a qualitative appreciation of uncertainty in a risk assessment outcome because we would like to be able to give quantitative estimates of the number of people at risk. This is where statistical methods that can capture uncertainty due to a lack of knowledge come into risk assessment. In particular, Bayesian statistical methods give us a framework for capturing both variability in populations and experimental results and uncertainty due to a lack of knowledge. There have been several examples

Table 2: The layout of an uncertainty table with some example entries.

Source of uncertainty	**Direction and magnitude**	**Supporting evidence**
Assumption that clinical studies are representative of general population (and description of impact on risk assessment endpoint)	$--/+$	Article explaining population variability
Assumption that physiological parameters derived from murine experiments are relevant (and description of impact on risk assessment endpoint)	$--/+$	Article explaining validity of assumption
Modelling assumption that the compartments are well-mixed with respect to chemical distribution (and description of impact on risk assessment endpoint)	$-/+$	Report showing effect of relaxing assumption
Assumption that the model is an accurate representation of reality (and description of impact on risk assessment endpoint)	$--/++$	Report covering validity of model choices
$\vdots$	$\vdots$	$\vdots$
Overall effect of identified uncertainties: Description of combined effect of preceding uncertainties.	$---/++$	

of Bayesian approaches being used in chemical risk assessment [47, 48], and [49] provides an overview of the Bayesian philosophy that is aimed at practitioners of computational toxicology. The key features of a Bayesian approach is the ability to combine many disparate sources of information and the flexibility to model uncertainty that occurs due to a lack of knowledge and inherent randomness simultaneously. Using Bayesian statistical modelling, we can move towards quantifying some of the uncertainties already mentioned in this section and help focus future research on areas where reduction of uncertainty will help risk assessors. In particular, we can calibrate the model in the light of the available data and reduce our uncertainty in the model parameters [47, 50, 51].

The ability to incorporate expert knowledge into an analysis makes Bayesian modelling even more appealing for chemical risk assessors because it allows them to formalise part of their typical assessment process. Expert elicitation is the process of deriving quantitative measures of experts' uncertainty [52, 53]. For our purpose, the elicitation exercise could tackle the uncertainty about a model parameter value [54], the likelihood of an event on an adverse outcome pathway [55] or the difference between some measure of true toxicity in humans as opposed to some other animal [56]. The methods and protocols that have been

established for the purpose of eliciting expert knowledge are far beyond simply asking experts for their best guess: they aim to minimise biases in judgements and to maintain transparency [details of modern protocols for expert elicitation are given in 57, 58].

Careful uncertainty quantification for chemical risk assessment is worthless if it is not communicated effectively to the decision maker. The results of the uncertainty analysis should be presented in a way such that the decision maker or risk manager can incorporate it into their (hopefully quantitative) decision analysis. Recently, there has been a great deal of research into effective communication of quantitative analyses and uncertainties [59, 60] and some discussion in the context of computational toxicology in [61]. Alongside these efforts, [62] have recently provided guidance on the communication of uncertainty in the risk assessment context.

4. Modelling in next generation risk assessment

Since the turn of the century, there has been significant effort in the development of non-animal approaches to chemical risk assessment. This effort has been driven by an ethical and regulatory need [63] and a desire to take advantage of recent technological advances in the study of human biology [64]. This is elevating the role of *in vitro* and *in silico* models in characterising the key biological events that lead to toxicity [7, 65, 66]. NGRA is an approach to chemical risk assessment that aims to integrate predictions from mathematical models with *in vitro* experiments that investigate biological, chemical and toxicological properties [6]. Indeed, future risk assessments should be able to combine mathematical models with relevant *in vitro* experiments to help risk assessors make predictions about adverse effects in a population of interest without the use of animals:

> "Advances in toxicogenomics, bioinformatics, systems biology, epigenetics, and computational toxicology could transform toxicity testing from a system based on whole-animal testing to one founded primarily on *in vitro* methods that evaluate changes in biologic processes using cells, cell lines, or cellular components, preferably of human origin" [67]

A key challenge is to understand how the different proposed elements of such a chemical risk assessment (for example, chemical characterisation and *in vitro* toxicity testing data) can be integrated to enable robust risk assessments. There are clear complexities due to the differing scales in the models and the multitude of uncertainties at each stage. Also, to incorporate it in a risk assessment, assessors need to understand the output accuracy of such an integrated framework of models and *in vitro* data.

Recent attempts to justify the adoption of non-animal alternatives (including *in vitro* systems and mathematical models) use statistical methods to evaluate the prediction accuracy of the new system by considering the level of correspondence with existing animal data. With such a strategy, the better the

goodness-of-fit, the better the predictor, and the mechanistic understanding is often ignored. There have also been some efforts to link the results of *in vitro* experiments and mathematical models to *in vivo* outcomes using Bayesian belief networks in [68] and influence diagrams in [56]. However, these approaches again largely ignore the detailed mechanistic understanding that, hopefully, underpins mathematical models of biological processes [as recommended in 23, 10]. Reverse dosimetry approaches [like those described in 69] directly link *in vitro* and exposure information via simple kinetics modelling. The uncertainties in such approaches due to the unknown link between the *in vitro* outcomes and human responses and the quality of the parameter estimates have led them to be screening tools for risk assessment rather than predictive tools for chemical potency [70]. Confidence in such approaches could be increased if there was break from overly simplistic modelling assumptions (like the link between the crude internal dose measurements and *in vitro* endpoints) and characterisation of the associated uncertainties.

Mechanistic models have the potential to be much more than extra strands in a risk assessor's weight-of-evidence approach. We can improve prediction by using a statistical framework that places mechanistic models at the centre of a toxicological risk assessment [71, 72]. By doing this, we enable experts to make judgements about how the elements of the mathematical model link various sources of experimental evidence and, more importantly, we create a system that focuses on the human end-points of interest (cancer incidence or adverse reactions in a consumer population for example). If adopted, these approaches could give a better prediction of potential effects in humans accounting for the inherent uncertainty, reduce the time and monetary costs of reliable risk assessment, and reduce the reliance on animals. A schematic of such a framework showing potential links and conditional independencies between the true population effects and external data is given in Figure 1. The key element in the framework is that the disparate data sources are all being used to inform the risk end-point of interest, which is crucial because we are not only following the principles of [7] by translating from models to meaningful risk end-points: we will have a framework within which to discuss the chances of adverse events occurring for our populations of interest.

In Figure 1, we show how the true human biological processes and true chemical properties have given rise to experimental data either directly, if the experiments are on humans or the chemical of interest, or indirectly, if the experiments are performed on another species, *in vitro* or on related chemicals. In theory, each of the biological-, physiological- and chemical-specific parameters of the mechanistic model has a true underlying value. By true, we are referring to what we would observe if we were able to take perfect measurements from the exact population of interest. In practice, we never know the true values due to limitations in our knowledge and the inability to test every individual in our population of interest. In such a framework, we must characterise, and aim to quantify, the variability in experimental results, the uncertainty in extrapolating from experiment to reality (especially when crossing from non-human or *in vitro* studies) and the variability in the population of interest. The variability

may be characterised using traditional statistical modelling methods, but extrapolation steps are more challenging as due to the unobservable nature of the parameters of interest in which case, as mentioned in Section 3, methods such as expert elicitation and model calibration will be needed. The distinction in the network between unobservable true states and experimental results allows the uncertainty regarding the adequacy of the assays to be separated from the uncertainty concerning interspecies differences between laboratory animals and humans and differences between animal systems and inherent properties of the chemical. This distinction was made in the method proposed by [56] with a difference that the true human response is easier to link to due to the mechanistic understanding within the mathematical model. Unlike in animal-based methods where interspecies differences are hopefully covered by safety factors or ignored, the multitude of uncertainties in the extrapolations can be handled within this framework. Discussions of the additional uncertainties that dealing with animal data bring can be found in [73, 74, 75, 76, 77].

The framework shown in Figure 1 focusses on the use of single models for each part of the process. We may have scenarios where the experts and model builders have several competing models for a biological process. Multiple models could be accommodating within this framework in two complementary ways. First, the uncertainty table approach, as described in Section 3, allows the experts to lay bare the strengths and weaknesses of the models, which could lead to some models being abandoned if the scientific basis is weak. Secondly, if several models were thought to be potentially useful, they could be accommodated within a hierarchical structure. The challenge here for the experts and the statistician would be to capture the interdependencies between the models given the possible shared scientific understandings being codified and the complications of common input parameters.

Although the focus has been on toxicity modelling thus far, we also have uncertainty about exposure levels of chemicals, and we can propagate this uncertainty about exposure model parameters whilst taking account of the variability in exposures in the population of interest. Mechanistic models for exposure are just as developed as toxicity models [78, 79] and some work has been done to account for variability in populations alongside the uncertainty for the model itself [80, 81]. Clearly, time can be saved with regards to formal quantitative uncertainty modelling if extreme estimates of tolerable dose and exposure are far apart. Properly accounting for uncertainty in chemical risk assessments can require a great deal of effort (although some of the results could be reused from one chemical assessment to the next); therefore, a tiered approach to the risk assessment is worthwhile [as suggested in both 6, 41].

5. Discussion

Despite investment and research in developing biologically-relevant mathematical models, there is not a general acceptance of the value of mathematical models in a risk assessment context. The challenge is getting the risk assessors to see the value of using models where, historically, chemical risk assessments

that have been driven by animal data. Part of this shift might be in the acceptance that animal models, *in vitro* models and mathematical models share many common features. One of the main aims of developing mathematical models is to reduce the need for animal use in toxicological research and risk assessments. However, it is false to think that mathematical models are not compatible with more traditional laboratory-based approaches: most mathematical models will be created using knowledge that has been established through laboratory experiments. Confidence in using mathematical models can also be improved by risk assessors embracing a benefit of accounting for uncertainty: uncertainty quantification can determine how conservative the risk management decisions are, and, typically, risk managers want to identify a safe level of chemical use without being unnecessarily over-conservative.

To boost confidence amongst risk assessors in using mathematical modelling for chemical risk assessments, there must be communication between the mathematical modellers and the risk assessors. The human body is complicated, and most realistic mechanistic models are also mathematically complicated. Mathematical and statistical models have a much better chance of acceptance if the model builders communicate regularly with the end users and are clear on how their modelling assumptions are affecting model behaviour. Software for building models and performing uncertainty analyses and probabilistic risk assessments are making mathematical models more accessible. Software for these purposes go from tailor-made solutions like Lhasa Limited's structural alert software Derek Nexus [82] and Certara's PBPK modelling platform [83] to programming languages like R and Python with libraries available to aid modelling and computation (like PySB [84]). There has also been some efforts to standardise modelling, coding and subsequent analyses that is aimed at modellers and toxicologists (for instance, the Systems Biology Markup Language [85]). In order to fully exploit the power of these methods, more effort is needed to educate risk assessors of the clear benefits, and the risk assessors need to work closely with modellers and statisticians to understand the utility of the computational methods and to make implementation smoother. Of course, expertise in toxicology, chemistry or biology does not immediately translate into expertise in computer science, mathematical modelling or statistical analysis.

As mentioned in Section 3, inevitably, there will be gaps in quantitative uncertainty analyses. In line with the principles of NGRA, the key to having a useful model that the experts and risk assessors have confidence in is transparency of the model-building process so that the users understand the limitations of the model and can make informed judgements about the potential deficiencies of the model. As such, by properly employing the models described herein and accounting for uncertainty, we can reduce the number of animals used in future studies (either to evaluate the toxicity of new chemicals or to help understand biological processes) and put historical experimental data to better use. This because an appreciation of uncertainty about a chemical risk end-point will help inform where further experiments might reduce uncertainty and what information we can take from existing studies.

Acknowledgements

Initial work for this article was supported by a grant from the UK's National Centre for Replacement, Refinement and Reduction of Animals in Research (grant NC/K001280/1).

References

[1] N. Burden, F. Sewell, K. Chapman, Testing chemical safety: what is needed to ensure the widespread application of non-animal approaches?, PLoS Biology 13 (5) (2015) e1002156.

[2] A. Paini, J. A. Leonard, T. Kliment, Y.-M. Tan, A. Worth, Investigating the state of physiologically based kinetic modelling practices and challenges associated with gaining regulatory acceptance of model applications, Regulatory Toxicology and Pharmacology 90 (2017) 104–115.

[3] J. C. Helton, Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal, Reliability Engineering & System Safety 42 (2-3) (1993) 327–367.

[4] H. de Moel, J. Aerts, Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates, Natural Hazards 58 (1) (2011) 407–425.

[5] J. P. Gosling, A. Hart, D. C. Mouat, M. Sabirovic, S. Scanlan, A. Simmons, Quantifying experts uncertainty about the future cost of exotic diseases, Risk Analysis 32 (5) (2012) 881–893.

[6] M. Dent, R. T. Amaral, P. A. Da Silva, J. Ansell, F. Boisleve, M. Hatao, A. Hirose, Y. Kasai, P. Kern, R. Kreiling, et al., Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients, Computational Toxicology 7 (2018) 20–26.

[7] G. T. Ankley, R. S. Bennett, R. J. Erickson, D. J. Hoff, M. W. Hornung, R. D. Johnson, D. R. Mount, J. W. Nichols, C. L. Russom, P. K. Schmieder, et al., Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment, Environmental Toxicology and Chemistry 29 (3) (2010) 730–741.

[8] D. Noble, T. J. Colatsky, A return to rational drug discovery: computer-based models of cells, organs and systems in drug target identification, Emerging Therapeutic Targets 4 (1) (2000) 39–49.

[9] J. Hoeng, M. Talikka, F. Martin, A. Sewer, X. Yang, A. Iskandar, W. K. Schlage, M. C. Peitsch, Case study: the role of mechanistic network models in systems toxicology, Drug discovery today 19 (2) (2014) 183–192.

[10] C. Wittwehr, H. Aladjov, G. Ankley, H. J. Byrne, J. de Knecht, E. Heinzle, G. Klambauer, B. Landesmann, M. Luijten, C. MacKay, et al., How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology, Toxicological Sciences 155 (2) (2017) 326–336.

[11] T. B. Knudsen, D. A. Keller, M. Sander, E. W. Carney, N. G. Doerrer, D. L. Eaton, S. C. Fitzpatrick, K. L. Hastings, D. L. Mendrick, R. R. Tice, et al., FutureTox II: in vitro data and in silico models for predictive toxicology, Toxicological Sciences 143 (2) (2015) 256–267.

[12] G. J. Myatt, E. Ahlberg, Y. Akahori, D. Allen, A. Amberg, L. T. Anger, A. Aptula, S. Auerbach, L. Beilke, P. Bellion, et al., In silico toxicology protocols, Regulatory Toxicology and Pharmacology 96 (2018) 1–17.

[13] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, B. P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling, Journal of Chemical Information and Computer Sciences 43 (6) (2003) 1947–1958.

[14] C. Zhao, H. Zhang, X. Zhang, M. Liu, Z. Hu, B. Fan, Application of support vector machine (SVM) for prediction toxic activity of different data sets, Toxicology 217 (2-3) (2006) 105–119.

[15] T. Luechtefeld, C. Rowlands, T. Hartung, Big-data and machine learning to revamp computational toxicology and its use in risk assessment, Toxicology Research.

[16] S. Kortagere, M. Lill, J. Kerrigan, Role of computational methods in pharmaceutical sciences, in: Computational Toxicology: Volume I, Springer, 2012, pp. 21–48.

[17] J. Jaworska, Y. Dancik, P. Kern, F. Gerberick, A. Natsch, Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice, Journal of Applied Toxicology 33 (11) (2013) 1353–1364.

[18] J. Devillers, Methods for building QSARs, in: Computational Toxicology, Springer, 2013, pp. 3–27.

[19] G. Patlewicz, G. Helman, P. Pradeep, I. Shah, Navigating through the minefield of read-across tools: A review of in silico tools for grouping, Computational Toxicology 3 (2017) 1–18.

[20] R. De Boer, P. Hogeweg, H. Dullens, R. A. De Weger, W. Den Otter, Macrophage T lymphocyte interactions in the anti-tumor immune response: a mathematical model., Journal of Immunology 134 (4) (1985) 2748–2758.

[21] M. Davies, R. U. Pendlington, L. Page, C. S. Roper, D. J. Sanders, C. Bourner, C. K. Pease, C. MacKay, Determining epidermal disposition kinetics for use in an integrated nonanimal approach to skin sensitization risk assessment, Toxicological Sciences 119 (2) (2010) 308–318.

[22] M. Yoon, G. L. Kedderis, G. Z. Yan, H. J. Clewell III, Use of in vitro data in developing a physiologically based pharmacokinetic model: carbaryl as a case study, Toxicology 332 (2015) 52–66.

[23] OECD, Guidance document on the validation of (quantitative) structure-activity relationships models, ENV/JM/MONO 2 (2007) 1–154.

[24] A. Paini, J. Leonard, E. Joossens, J. Bessems, A. Desalegn, J. Dorne, J. Gosling, M. Heringa, M. Klaric, T. Kliment, et al., Next generation physiologically based kinetic (NG-PBK) models in support of regulatory decision making, Computational Toxicology.

[25] C. M. Ellison, S. J. Enoch, M. T. Cronin, A review of the use of in silico methods to predict the chemistry of molecular initiating events related to drug toxicity, Expert opinion on drug metabolism & toxicology 7 (12) (2011) 1481–1495.

[26] M. T. Cronin, A.-N. Richarz, Relationship between adverse outcome pathways and chemistry-based in silico models to predict toxicity, Applied In Vitro Toxicology 3 (4) (2017) 286–297.

[27] M. D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (2) (1991) 161–174.

[28] A. Boukouvalas, J. P. Gosling, H. Maruri-Aguilar, An efficient screening method for computer experiments, Technometrics 56 (2014) 422–31.

[29] A. Saltelli, K. Chan, E. Scott (Eds.), Sensitivity Analysis, New York: Wiley, 2000.

[30] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation 55 (1) (2001) 271–280.

[31] K. McNally, R. Cotton, G. D. Loizou, A workflow for global sensitivity analysis of PBPK models, Frontiers in Pharmacology 2 (2011) 31.

[32] A. Kruizinga, D. Briggs, R. Crevel, A. Knulst, L. Van den Bosch, G. Houben, Probabilistic risk assessment model for allergens in food: sensitivity analysis of the minimum eliciting dose and food consumption, Food and Chemical Toxicology 46 (5) (2008) 1437–1443.

[33] A. Kiparissides, S. Kucherenko, A. Mantalaris, E. Pistikopoulos, Global sensitivity analysis challenges in biological systems modeling, Industrial & Engineering Chemistry Research 48 (15) (2009) 7168–7180.

[34] J. Bernoulli, Ars Conjectandi, Impensis Thurnisiorum, Fratrum, 1713.

[35] Codex Alimentarius Commission, Appendix IV: Working principles for risk analysis for application in the framework of the Codex Alimentarius, in: Report of the Twenty-sixth Session, FAO Headquarters, Rome, 30 June to 7 July 2003, 2003.

[36] M. F. Festing, P. Overend, R. Gaines Das, M. Cortina-Borja, M. Berdoy, The design of animal experiments: reducing the use of animals in research through better experimental design, Vol. 1, The Royal Society of Medicine Press Limited, 2002.

[37] C. Kilkenny, W. J. Browne, I. C. Cuthill, M. Emerson, D. G. Altman, Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research, Animals 4 (1) (2014) 35–44.

[38] G. E. Dinse, J. K. Haseman, Logistic regression analysis of incidental-tumor data from animal carcinogenicity experiments, Toxicological Sciences 6 (1) (1986) 44–52.

[39] Y. Ren, H. Liu, C. Xue, X. Yao, M. Liu, B. Fan, Classification study of skin sensitizers based on support vector machine and linear discriminant analysis, Analytica Chimica Acta 572 (2) (2006) 272–282.

[40] J. F. Rathman, C. Yang, H. Zhou, Dempster-shafer theory for combining in silico evidence and estimating uncertainty in chemical risk assessment, Computational Toxicology 6 (2018) 16–31.

[41] EFSA, Guidance on uncertainty analysis in scientific assessments, EFSA Journal 16 (1).

[42] H. A. Barton, W. A. Chiu, R. Woodrow Setzer, M. E. Andersen, A. J. Bailer, F. Y. Bois, R. S. DeWoskin, S. Hays, G. Johanson, N. Jones, et al., Characterizing uncertainty and variability in physiologically based pharmacokinetic models: state of the science and needs for research and implementation, Toxicological Sciences 99 (2) (2007) 395–402.

[43] E. A. Patterson, M. P. Whelan, A framework to establish credibility of computational models in biology, Progress in Biophysics and Molecular Biology 129 (2017) 13–19.

[44] A. Hart, J. Gosling, A. Boobis, D. Coggon, P. Craig, D. Jones, Development of a framework for evaluation and expression of uncertainties in hazard and risk assessment, Final Report of Food Standards Agency Project (T01056).

[45] D. Tennant, J. P. Gosling, Modelling consumer intakes of vegetable oils and fats, Food Additives & Contaminants: Part A 32 (9) (2015) 1397–1405.

[46] EFSA, The principles and methods behind EFSA's guidance on uncertainty analysis in scientific assessment, EFSA Journal 16 (1).

[47] A. Gelman, F. Bois, J. Jiang, Physiological pharmacokinetic analysis using population modeling and informative prior distributions, Journal of the American Statistical Association 91 (436) (1996) 1400–1412.

[48] E. P. Grist, A. O'Hagan, M. Crane, N. Sorokin, I. Sims, P. Whitehouse, Bayesian and time-independent species sensitivity distributions for risk assessment of chemicals, Environmental Science & Technology 40 (1) (2006) 395–401.

[49] F. Y. Bois, Computational Toxicology: Volume II, Humana Press, New York, 2013, Ch. Bayesian Inference, pp. 596–636.

[50] D. Henderson, R. Boys, D. Wilkinson, Bayesian calibration of a stochastic kinetic computer model using multiple data sources, Biometrics 66 (1) (2010) 249–256.

[51] A. Hawkins-Daarud, S. Prudhomme, K. G. van der Zee, J. T. Oden, Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth, Journal of Mathematical Biology 67 (6-7) (2013) 1457–1485.

[52] R. M. Cooke, Experts in Uncertainty: Opinion and Subjective Probability in Science, Oxford University Press, New York, 1991.

[53] A. O'Hagan, C. E. Buck, A. Daneshkhah, J. E. Eiser, P. H. Garthwaite, D. Jenkinson, J. E. Oakley, T. Rakow, Uncertain judgements: eliciting expert probabilities, Wiley, Chichester, 2006.

[54] H. Van der Fels-Klerx, R. M. Cooke, M. N. Nauta, L. H. Goossens, A. H. Havelaar, A structured expert judgment study for a model of campylobacter transmission during broiler-chicken processing, Risk Analysis 25 (1) (2005) 109–124.

[55] A. B. Knol, J. J. de Hartog, H. Boogaard, P. Slottje, J. P. van der Sluijs, E. Lebret, F. R. Cassee, J. A. Wardekker, J. G. Ayres, P. J. Borm, et al., Expert elicitation on ultrafine particles: likelihood of health effects and causal pathways, Particle and Fibre Toxicology 6 (1) (2009) 19.

[56] J. P. Gosling, A. Hart, H. Owen, M. Davies, J. Li, C. MacKay, et al., A Bayes linear approach to weight-of-evidence risk assessment for skin allergy, Bayesian Analysis 8 (1) (2013) 169–186.

[57] J. P. Gosling, SHELF: the Sheffield elicitation framework, in: Elicitation: the science and art of structuring judgement, Springer, 2018, pp. 61–93.

[58] J. Quigley, A. Colson, W. Aspinall, R. M. Cooke, Elicitation in the classical model, in: Elicitation: the science and art of structuring judgement, Springer, 2018, pp. 15–36.

[59] B. Fischhoff, A. L. Davis, Communicating scientific uncertainty, Proceedings of the National Academy of Sciences 111 (Supplement 4) (2014) 13664–13671.

[60] J. Roberts, P. Gough, Communicating statistical uncertainty to non-expert audiences: Interactive disease mapping, in: Big Data Visual Analytics (BDVA), 2016, IEEE, 2016, pp. 1–3.

[61] W. Mendez, B. A. Fowler, Translation of computational model results for risk decisions, in: Computational Toxicology, Elsevier, 2013, pp. 215–245.

[62] EFSA, Clear communications and uncertainty: External scientific report. RC/EFSA/COMM/2015/01, EFSA Supporting Publications 15 (5) (2018) 1412E.

[63] S. Adler, D. Basketter, S. Creton, O. Pelkonen, J. Van Benthem, V. Zuang, K. E. Andersen, A. Angers-Loustau, A. Aptula, A. Bal-Price, et al., Alternative (non-animal) methods for cosmetics testing: current status and future prospects, Archives of Toxicology 85 (5) (2011) 367–485.

[64] D. Krewski, D. Acosta Jr, M. Andersen, H. Anderson, J. C. Bailar III, K. Boekelheide, R. Brent, G. Charnley, V. G. Cheung, S. Green Jr, et al., Toxicity testing in the 21st century: a vision and a strategy, Journal of Toxicology and Environmental Health, Part B 13 (2-4) (2010) 51–138.

[65] Y. Adeleye, M. Andersen, R. Clewell, M. Davies, M. Dent, S. Edwards, P. Fowler, S. Malcomber, B. Nicol, A. Scott, et al., Implementing Toxicity Testing in the 21st Century (TT21C): Making safety decisions using toxicity pathways, and progress in a prototype risk assessment, Toxicology 332 (2015) 102–111.

[66] S. W. Edwards, Y.-M. Tan, D. L. Villeneuve, M. Meek, C. A. McQueen, Adverse Outcome Pathways: Organizing toxicological information to improve decision making, Journal of Pharmacology and Experimental Therapeutics 356 (1) (2016) 170–181.

[67] National Research Council, Toxicity testing in the 21st century: a vision and a strategy, National Academies Press, 2007.

[68] J. Jaworska, A. Harol, P. S. Kern, G. F. Gerberick, Integrating non-animal test information into an adaptive testing strategy: skin sensitization proof of concept case., ALTEX 28 (2011) 211–25.

[69] J. Louisse, K. Beekmann, I. M. Rietjens, Use of physiologically based kinetic modeling-based reverse dosimetry to predict in vivo toxicity from in vitro data, Chemical Research in Toxicology 30 (1) (2016) 114–125.

[70] B. A. Wetmore, J. F. Wambaugh, B. Allen, S. S. Ferguson, M. A. Sochaski, R. W. Setzer, K. A. Houck, C. L. Strope, K. Cantwell, R. S. Judson,

et al., Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing, Toxicological Sciences 148 (1) (2015) 121–136.

[71] O. Taboureau, K. Audouze, Human environmental disease network: A computational model to assess toxicology of contaminants, ALTEX-Alternatives to animal experimentation 34 (2) (2017) 289–300.

[72] J. Reynolds, C. MacKay, N. Gilmour, D. Miguel-Vilumbrales, G. Maxwell, Probabilistic prediction of human skin sensitiser potency for use in next generation risk assessment, Computational Toxicology.

[73] L. S. Gold, T. H. Slone, B. N. Ames, What do animal cancer tests tell us about human cancer risk?: Overview of analyses of the carcinogenic potency database, Drug Metabolism Reviews 30 (2) (1998) 359–404.

[74] A. Knight, Animal experiments scrutinised: systematic reviews demonstrate poor human clinical and toxicological utility, ALTEX-Alternatives to animal experimentation 24 (4) (2007) 320–325.

[75] N. Shanks, R. Greek, J. Greek, Are animal models predictive for humans?, Philosophy, ethics, and humanities in medicine 4 (1) (2009) 2.

[76] I. W. Mak, N. Evaniew, M. Ghert, Lost in translation: animal models and clinical trials in cancer treatment, American journal of translational research 6 (2) (2014) 114.

[77] A. Akhtar, The flaws and human harms of animal experimentation, Cambridge Quarterly of Healthcare Ethics 24 (4) (2015) 407–419.

[78] J. de Boer, E. Fritsche, G. Schoeters, I. Kimber, The European Long-range Research Initiative (LRI): a decade of contributions to human health protection, exposure modelling and environmental integrity, Toxicology 337 (2015) 83–90.

[79] B. Safford, A. Api, C. Barratt, D. Comiskey, G. Ellis, C. McNamara, C. O'Mahony, S. Robison, J. Rose, B. Smith, et al., Application of the expanded Creme RIFM consumer exposure model to fragrance ingredients in cosmetic, personal care and air care products, Regulatory Toxicology and Pharmacology 86 (2017) 148–156.

[80] K. McNally, R. Cotton, A. Hogg, G. Loizou, PopGen: a virtual human population generator, Toxicology 315 (2014) 70–85.

[81] S. Kettler, M. Kennedy, C. McNamara, R. Oberdörfer, C. O'Mahony, J. Schnabel, B. Smith, C. Sprong, R. Faludi, D. Tennant, Assessing and reporting uncertainties in dietary exposure analysis: mapping of uncertainties in a tiered approach, Food and Chemical Toxicology 82 (2015) 79–95.

[82] C. A. Marchant, Computational toxicology: a tool for all industries, Wiley Interdisciplinary Reviews: Computational Molecular Science 2 (3) (2012) 424–434.

[83] Certara, Phoenix 8.1 (2018).
URL https://www.certara.com/software/pkpd-modeling-and-simulation-2/phoenix-8/

[84] C. F. Lopez, J. L. Muhlich, J. A. Bachman, P. K. Sorger, Programming biological models in Python using PySB, Molecular systems biology 9 (1) (2013) 646.

[85] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, et al., The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, Bioinformatics 19 (4) (2003) 524–531.

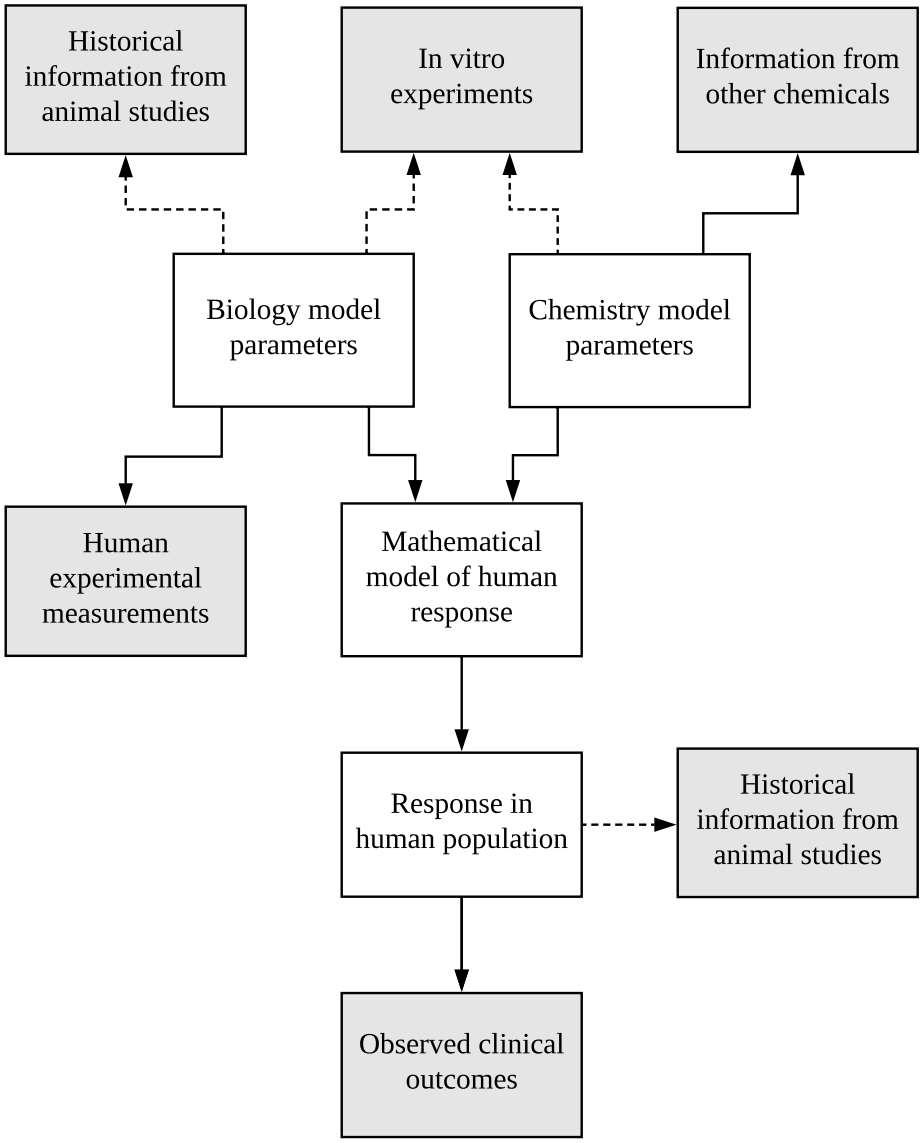

Figure 1: A network demonstrating how experimental results and observed data (given in shaded boxes) are determined by the true world situation (given in the unshaded boxes). Arrows that are dashed indicate that the translation from experiment to human relevant information would require an extrapolation across species or from *in vitro* to *in vivo*. On a technical note, the direction of the arrows give our judgements about conditional independence as could be derived from a standard Bayesian network.