"The conception of chance enters in the very first steps of scientific activity in virtue of the fact that no observation is absolutely correct. I think chance is a more fundamental conception than causality; for whether in a concrete case, a cause-effect relation holds or not can only be judged by applying the laws of chance to the observation."

Max Born (German-British theoretical physicist, 1882-1970)

t4 Workshop Report*

From Cellular Perturbation to Probabilistic Risk Assessments

Alexandra Maertens[1], Breanne Kincaid[1], Eric Bridgeford[2], Celine Brochot[3], Arthur de Carvalho e Silva[4], Jean-Lou C. M. Dorne[5], Liesbet Geris[6,7], Trine Husøy[8], Nicole Kleinstreuer[9], Luiz C. M. Ladeira[6], Alistair Middleton[10], Joe Reynolds[10], Blanca Rodriguez[11], Erwin L. Roggen[12], Giulia Russo[13], Kris Thayer[14] and Thomas Hartung[1,15,16]

[1]Center for Alternatives to Animal Testing (CAAT), Johns Hopkins Bloomberg School of Public Health and Whiting School of Engineering, Baltimore, MD, USA; [2]Dept. of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA; [3]Certara, Canterbury, Kent, UK; [4]University of Birmingham and Michabo Health Science Ltd, Birmingham, UK; [5]European Food Safety Authority (EFSA), Parma, Italy; [6]Molecular and Computational Biology, University of Liège, Liège, Belgium; [7]Mechanical Engineering, KU Leuven, Leuven, Belgium; [8]Food Safety/ Centre for Sustainable Diets, Norwegian Institute of Public Health (NIPH), Oslo, Norway; [9]National Institutes of Health, National Institute for Environmental Health Sciences, DTT/NICEATM, Durham, NC, USA; [10]Safety & Environmental Assurance Centre, Unilever, Colworth Science Park, Sharnbrook, Bedfordshire, UK; [11]Computer Science, University of Oxford, Oxford, UK; [12]3RsMC ApS, Kongens Lyngby, Lyngby, Denmark; [13]Drug and Health Sciences, University of Catania, Catania, Italy; [14]Chemical & Pollutant Assessment Division (CPAD), US Environmental Protection Agency (EPA), Durham, NC, USA; [15]Doerenkamp-Zbinden Chair for Evidence-based Toxicology, Johns Hopkins Bloomberg School of Public Health and Whiting School of Engineering, Baltimore, MD, USA; [16]CAAT-Europe, University of Konstanz, Konstanz, Germany

Received January 29, 2025;

Correspondence:
Thomas Hartung, MD, PhD,
Center for Alternatives to Animal Testing (CAAT),
Johns Hopkins University,
615 N Wolfe St., Baltimore, MD, 21205, USA
(thartun1@jhu.edu)

ALTEX 42(3), 413-434.
doi:10.14573/altex.2501291

Abstract

Chemical risk assessment is evolving from traditional deterministic approaches to embrace probabilistic methodologies, where risk of hazard manifestation is understood as a more or less probable event depending on exposure, individual factors, and stochastic processes. This is driven by advancements in human stem cells, complex tissue engineering, high-performance computing, and cheminformatics, and is more recently facilitated by large-scale artificial intelligence models. These innovations enable a more nuanced understanding of chemical hazards, capturing the complexity of biological responses and variability within populations. However, each technology comes with its own uncertainties impacting on the estimation of hazard probabilities. This shift addresses the limitations of point estimates and thresholds that oversimplify hazard assessment, allowing for the integration of kinetic variability and uncertainty metrics into risk models. By leveraging modern technologies and expansive toxicological data, probabilistic approaches offer a comprehensive evaluation of chemical safety. This paper summarizes a workshop held in 2023 and discusses the technological and data-driven enablers, and the challenges faced in their implementation, with particular focus on perturbation of biology as the basis of hazard estimates. The future of toxicological risk assessment lies in the successful integration of these probabilistic models, promising more accurate and holistic hazard evaluations.

Plain language summary

Understanding chemical risks is key to public health. Traditional risk assessments rely on fixed safety margins and animal tests, which can miss complex human responses. Probabilistic risk assessment uses advanced tools – human stem cells, organ-on-chip systems, and AI – to estimate the likelihood of harm across different scenarios. By modeling individual variability (genetics, exposures) and quantifying uncertainty, it provides nuanced risk estimates rather than binary "safe/unsafe" labels. This approach increases transparency, shows confidence intervals, and reduces animal testing by integrating human-relevant data. Challenges include defining harm thresholds, integrating diverse datasets, and gaining regulatory acceptance. Workshops like the 2023 CAAT-ONTOX meeting in Italy highlighted how measuring biological perturbations (e.g., molecular or cellular changes) informs probability of adverse outcomes. As technologies and data improve, probabilistic methods promise more realistic, protective chemical safety evaluations that reflect real-world human diversity.

*A report of t4 – the transatlantic think tank for toxicology, a collaboration of the toxicologically oriented chairs in Baltimore and Konstanz sponsored by the Doerenkamp-Zbinden Foundation. The spectrum of views expressed in this article are those of the contributing authors and do not necessarily reflect those of their institution of employment.

1 Introduction

Chemical risk assessment is in the midst of the paradigm shift sweeping the field of toxicology: With great advancements in human stem cells, complex tissue engineering, high-performance computing, cheminformatics, and large-scale artificial intelligence (AI) models, the toxicological sciences are better poised now than at any point in the field's history to predict chemical hazard. We can characterize in detail how organisms react to chemical exposure, but the challenge is to translate this data to an individual's and population's risk to enable risk management. This requires the integration of very diverse information from these advanced technologies and calls for the adoption of probabilistic approaches to hazard assessment (Maertens et al., 2022).

The European ONTOX project[1] (Vinken et al., 2021) aims at promoting safer chemicals using fewer animals and places probabilistic risk assessment (ProbRA) at its core to enhance the accuracy and reliability of hazard predictions and risk characterizations. ProbRA integrates diverse data sources and models the variability and uncertainty inherent in biological responses to chemical exposures. By employing systems toxicology approaches, ONTOX leverages high-throughput omics data and computational modeling to identify and quantify the probability of adverse outcomes.

Traditional risk assessment methods often rely on deterministic approaches, which use fixed safety margins and thresholds to evaluate chemical hazards. These are often based on propagating forward numerous worst-case scenario assumptions, which produce conservative point estimates. However, these methods do not adequately capture the variability and uncertainty inherent in biological systems. Systems toxicology offers a probabilistic approach to risk assessment, which accounts for the distribution of responses within a population and the uncertainty associated with different data sources. Probabilistic hazard assessment uses statistical models to estimate the likelihood of adverse outcomes based on the extent of biological perturbation. This approach allows for a more nuanced understanding of risk, incorporating factors such as dose-response relationships, inter-individual variability, and the combined effects of multiple pathways.

At the heart of this evolution lies a critical question: Why do we require a departure from deterministic approaches towards embracing probabilistic methodologies in hazard assessment? The assumption of a binary presence of hazard at a given dose falls short of capturing the biological interplay between adaptive response and perturbation. Hierarchical classification systems describing the strength of hazard evidence are the typical regulatory output of deterministic hazard assessment, with a primary example being the World Health Organization's (WHO's) International Agency for Research on Cancer (IARC), which groups chemical agents from Group 1 ("carcinogenic to humans") through Group 2A ("Probably carcinogenic to humans"), 2B ("Possibly carcinogenic to humans"), and 3 ("Not classifiable as to its carcinogenicity to humans"). Moreover, the inherent variability within experimental models and human populations – shaped by factors ranging from nutritional status and co-exposures to genetic backgrounds – diminishes the informative value of deterministic approaches. In terms of setting enforceable contaminant levels, conventional efforts have ignored this variability altogether, applying arbitrary uncertainty factors (i.e., 10-fold dose reductions) to extrapolate to the population from the experiment.

This discussion of variability and uncertainty highlights one of the many benefits of adopting probabilistic approaches in hazard assessment: empowering risk assessors to embed concrete metrics of uncertainty into their models. Models can now incorporate kinetic variability, delineating hazard characterization through points of departure by internal dose, target organ, or even subcellular targets. As an additional layer of confidence, probabilistic models that capture and enumerate any measurable variability and the uncertainty that accompanies it can be constructed. This biologically informed perspective enables a more holistic integration of spatial and temporal variability, fostering a systems-level comprehension of hazard.

The transition to ProbRA is made possible by modern technologies and the exponential growth of available toxicological data and computing power. Text mining of peer-reviewed and grey literature allows for the capture of a broader array of relevant data than traditional (systematic) reviews (Hoffmann et al., 2017) and meta-analyses, informing a more comprehensive weight-of-evidence approach. Furthermore, the application of AI helps overcome limited understanding of biological interactions under a systems biology perspective (Hartung et al., 2017), enabling computers to identify patterns in large datasets that may not be obvious to human observers.

However, this promising future also presents its own challenges, particularly as the current hazard assessment landscape shifts towards greater adoption of *in vitro* and *in silico* models. This shift necessitates distinguishing between adaptive responses and adverse outcomes (Hartung, 2023a). For example, mere receptor binding or gene expression changes do not necessarily indicate a negative impact on functionality, as biological systems are inherently adaptive. Additionally, challenges arise in defining adversity using highly dimensional data, such as omics data, which is inherently upstream of the phenotypic manifestation of hazard. Determining the minimum acceptable evidence of an adverse outcome based on differentially expressed genes, translated proteins, generated metabolites, or altered subcellular functionality remains a complex task.

[1] https://ontox-project.eu/

Abbreviations: AI, artificial intelligence; AOP, adverse outcome pathway; ECHA, European Chemicals Agency; EFSA, European Food Safety Authority; IARC, International Agency for Research on Cancer; ProbRA, probabilistic risk assessment; KE, key event; MIE, molecular initiating event; MPS, microphysiological systems; NAM, new approach methodology; OECD, Organisation for Economic Co-operation and Development; PBPK, physiologically-based pharmacokinetic; POD, point of departure; (Q)IVIVE, quantitative *in-vitro*-to-*in-vivo* extrapolation; WoE, weight of evidence

The discussions at the workshop organized by the Center for Alternatives to Animal Testing (CAAT) and ONTOX, financed by the Doerenkamp-Zbinden Foundation as part of the t[4] Transatlantic Think Tank for Toxicology, in Angera, Italy from 4 to 6 July 2023, centered around the data sources and technologies that enable this transition to ProbRA, as well as the applications and challenges associated with this shift. In particular, the workshop addressed how the perturbation of biological processes can contribute to the assessment of the probability of hazardous effects beyond the chemicals' structural components and test data. By addressing these key aspects, the workshop aimed to provide insights into the future of toxicological risk assessment and its potential to leverage modern technologies and data sources for more accurate and comprehensive evaluations of chemical safety.

2 The current state of probabilistic risk assessment

The field of toxicology and risk assessment is undergoing a significant transformation, driven by the need for more accurate, efficient, and human-relevant approaches. Despite the substantial investment into safety assessments, with annual spending reaching $20 billion (Meigs et al., 2018), used to regulate multi-trillions of marketed products, the primary methods used in regulatory toxicology were introduced 50 and more years ago. The reliance on animal toxicity testing has been a cornerstone of risk assessment practices, yet its ability to predict human side effects and interspecies differences remains limited. The thalidomide tragedy serves as a classic example of how interspecies differences in phocomelia hazard allowed a drug with a critical period of hazard to be introduced to the market.

Cheminformatic models, which predict toxicity based on chemical structure, can achieve accuracy levels > 80%, depending on the hazard endpoint (Luechtefeld et al., 2016a, 2018a; Golden et al., 2023). This provides the opportunity to integrate aspects of biology to bridge activity cliffs and further improve predictivity. It is becoming increasingly clear that treating small mammals as a substitute for humans is no longer acceptable, yet chemical toxicity test guidelines still uphold animal use as the gold standard.

Probabilistic modeling, which has typically been limited to exposure assessment, is gaining traction in the field of risk assessment (Maertens et al., 2022; Golden et al., 2024). Examples of current regulatory use, available tools, and the confidence in assessments demonstrate the potential of probabilistic approaches. However, it is essential to recognize that risk assessment is, by nature, a prediction, and the ability to integrate diverse data sources and technologies is crucial for its advancement.

2.1 Data sources

The amount of available toxicological data is increasing continuously. Moreover, the diversity of data has increased, necessitating the incorporation of non-guideline studies, *in vitro* technologies, high-throughput and high-dimensional data such as omics, and *in chemico* approaches into ProbRAs. The challenge lies not in the scarcity of data-rich sources but in the ability to efficiently and transparently integrate these disparate data. While the necessary computing power is now available to allow the integration of high-dimensional data across different data structures, this will only be useful to regulators if there is confidence in the method of data integration. Larger datasets do not guarantee that the signal will be stronger than the noise. Factors such as context of use, biological mechanism, incorporation of quality controls, and independent review must be considered. Furthermore, publication bias in toxicology tends to favor positive results, while most chemicals in the European Chemicals Agency (ECHA) registration database are negative (Luechtefeld et al., 2016b), emphasizing the need for unsupervised data gathering efforts. Going forward, the scientific community needs to valorize negative results. The concept of information economy becomes relevant, requiring a precise understanding of the best test to reduce uncertainty at the lowest cost (value of information). Some cases will be exposure-driven, while others will be mechanistically driven.

2.2 Technologies

Advancements in computational methodologies, such as big data approaches, the ability to scoop legacy data, and the accessibility of high-performance computing, are driving the evolution of ProbRA. The ToxCast program[2], which utilizes *in vitro* technologies, has proven useful for bio-based read-across, although significant data gaps and challenges remain: ToxCast is largely based on cancer-derived cell lines, and the *in vitro* data has been built in a piecemeal manner over time, with challenges to ensure the necessary quality control for confidence in risk assessment. Microphysiological systems (MPS) (Marx et al., 2016, 2020, 2025) may serve as a potential bridge (Hartung and Smirnova, 2025). High-performance computing capabilities have opened up new possibilities, such as molecular docking and molecular dynamics simulations, which were previously inaccessible for risk assessment. The growing diversity of human data, including omics, large datasets like The Cancer Genome Atlas[3], electronic health records, and new biomonitoring sources, presents opportunities for data mining and integration. Natural language processing (NLP) systems are advanced enough to mine molecular interactions from the literature and monitor new publications in real time with confidence (Gyori et al., 2017; Bachman et al., 2023). Applications to toxicological data are also under development (Hartung, 2023b,c; Kleinstreuer and Hartung, 2024; Corradi et al., 2024). AI techniques can provide better readouts of cell health and phenotype compared to specific molecular readouts or receptor binding, with downstream effects being better captured by AI, while docking, further refined by molecular dynamics simulations, may be more suitable for specific receptor binding.

To fully harness the potential of ProbRA, several key applications must be developed. These include:

2 https://www.epa.gov/comptox-tools/toxicity-forecasting-toxcast

3 https://www.cancer.gov/ccg/research/genome-sequencing/tcga

1. Predictive toxicology tools such as read-across-based structure activity relationships (RASAR) (Hartung, 2016; Luechtefeld and Hartung, 2017; Luechtefeld et al., 2018a,b);
2. Physiological maps and quantitative adverse outcome pathways (AOPs)(Vinken et al., 2021);
3. Mapping molecular initiating events (MIEs) to probabilistic outcomes of individual and population health;
4. Ontologies for organizing and integrating knowledge;
5. Points of departure (PODs) based on internal dose, target organ, or subcellular targets;
6. BioBricks (Gao et al., 2024) for modular and reusable biological components;
7. Open-source human metabolite prediction tools, such as BioTransformer[4] and GLORYx (de Bruyn-Kops et al., 2021), which use biological rule-based models to identify novel biotransformation products from parent chemical structures and provide likelihood outputs for each prediction (Djoumbou-Feunang et al., 2019);
8. AI-guided tools like DeepDock (Liao et al., 2019), which integrate publicly available software such as Vina AutoDock[5], enabling bulk target receptor binding assessment of extensive chemical libraries, a task that would have been computationally prohibitive a decade ago (Alhossary et al., 2015).

In conclusion, the current state of ProbRA in toxicology is characterized by a growing recognition of the need for more human-relevant and predictive approaches, the availability of diverse data sources and advanced technologies, and the identification of key applications that can revolutionize the field. By addressing the challenges associated with data integration, quality control, and regulatory acceptance, the toxicology community can harness the full potential of ProbRA to improve public health and safety.

3 Advantages of probabilistic risk assessment

ProbRA offers numerous advantages over traditional deterministic approaches in toxicology and risk assessment. These include:

1. *Transparency and credibility:* ProbRA improves transparency and credibility by explicitly considering and treating all types of uncertainties in a structured, integrative, and quantitative manner. This allows for ranking of issues and results, providing more information by separating variability from uncertainty.
2. *Cost effectiveness:* By focusing resources on essential safety issues, ProbRA can be cost-effective. It helps prioritize and focus data collection efforts on parameters that have the greatest impact on risk estimates.
3. *More realism compared to deterministic approaches:* ProbRA avoids worst-case assumptions and provides more realistic exposure assessments. It gives an overall picture of risks in the population, not just extreme cases. A probabilistic reference dose could help reduce the potentially inaccurate implication of zero risk below the reference dose.
4. *Improved decision support:* ProbRA enables risk managers to evaluate the full range of variability and uncertainty instead of relying on conservative point estimates. This leads to more informed and robust decision-making.
5. *Systematic sensitivity analysis:* ProbRA includes a systematic sensitivity analysis of the uncertainties in the input parameters, identifying the main sources of uncertainty. This helps focus further research and data collection efforts.
6. *Optimization process:* ProbRA allows for the application of an optimization process, leading to more efficient and targeted risk assessment and management.
7. *More effective risk management:* By providing a fuller understanding of risks, ProbRA enhances safety and helps manage operability. It improves the estimation of the success of risk mitigation measures.
8. *Transparent risk communication:* ProbRA results and decisions can be communicated on a clearly defined basis, promoting transparency and trust in the risk assessment process.
9. *Functionality with limited data:* Even with relatively small amounts of data, ProbRA can be used as a comparative tool to make decisions between different design or operation alternatives. The absolute accuracy of the data is less critical in this context.
10. *Information economy:* ProbRA enables formally estimating the value of gathering additional information, allowing better prioritization of research efforts by investing in areas that yield the greatest information value.

In summary, ProbRA promotes transparency, credibility, realism, and cost-effectiveness in risk assessment (Fig. 1). It supports robust decision-making and risk communication by systematically quantifying uncertainties and variabilities. Even with limited data, ProbRA can provide comparative insights, and it helps prioritize further research. These advantages position ProbRA as a key tool for advancing toxicology and risk assessment in line with modern scientific understanding and societal expectations.

4 Challenges of probabilistic risk assessment

While ProbRA offers numerous advantages over traditional deterministic approaches, it also presents several challenges that must be addressed for its successful implementation and acceptance:

1. *Model incompleteness:* ProbRA models may suffer from incompleteness, which can be more apparent than in deterministic models. This "*might be a fertile ground for endless debate between utility and regulator*" (Kafka, 1998), leading to potential delays in regulatory decision-making.
2. *Complexity and time-consumption:* ProbRA involves a more complex and time-consuming analysis and decision-making process compared to deterministic approaches. It requires more data, as distributions of values rather than single values are used, and more information and insights must be collected, processed, and considered for decisions.

[4] https://biotransformer.ca

[5] https://www.computabio.com/protein-protein-docking-service.html

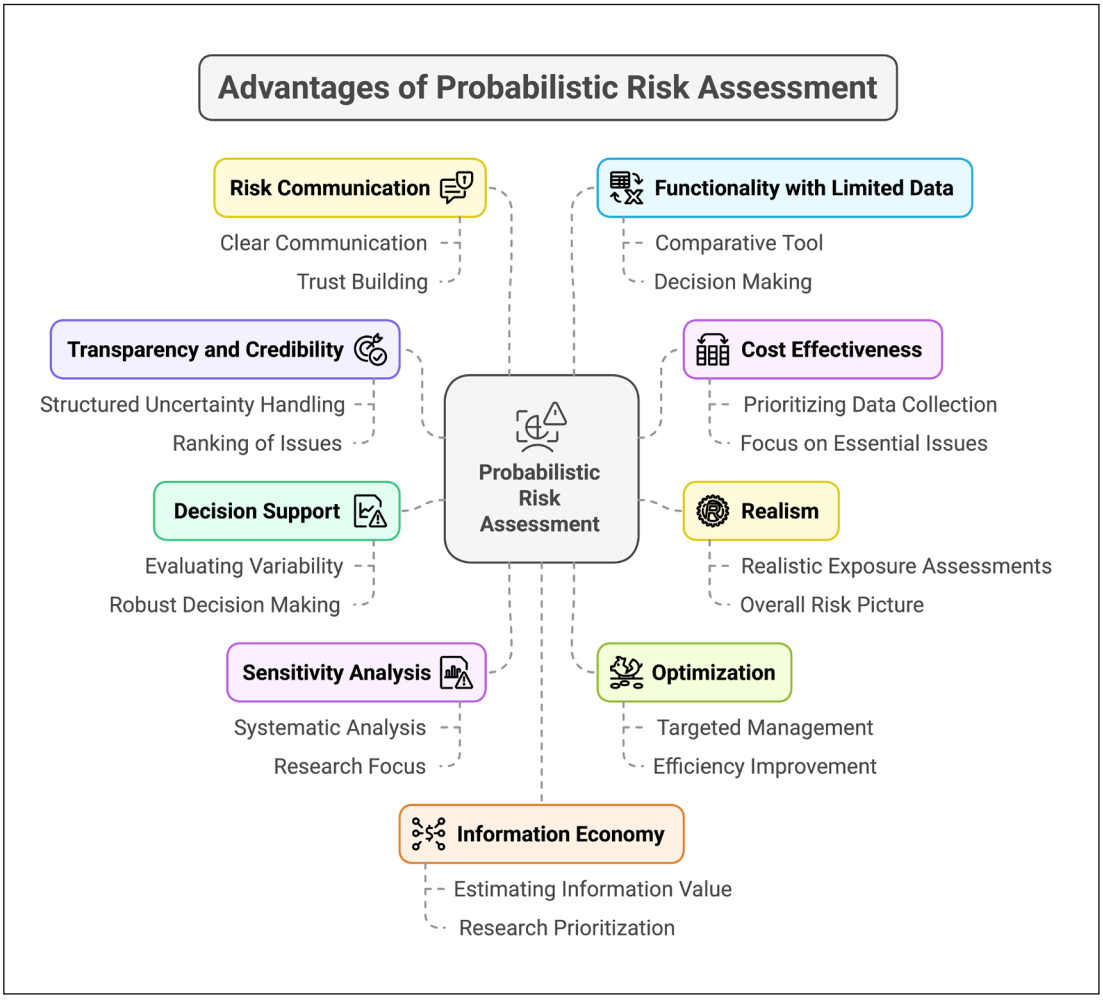

Fig. 1: Advantages of probabilistic risk assessment Designed with napkin.ai.

3. *Mathematical education:* The complex structure of ProbRA models and the associated assumptions, methods, and results can be difficult to understand for those without a strong mathematical background. This may require additional training for personnel to ensure consistent application of standards.
4. *Extremely rare events:* When considering extremely rare events, ProbRA may face challenges in ensuring the statistical significance of probabilistic data. This can limit the reliability of risk estimates for such events.
5. *Validation and good practices:* Validating ProbRA models can be challenging, as it may be unclear what to compare the results against. Additionally, there is currently a lack of established good practices for ProbRA in toxicology, which can hinder its consistent application.
6. *Decision-making complications:* In some cases, the more comprehensive characterization of uncertainties in ProbRA may lead to a decrease in clarity regarding how to estimate risk for the scenario under consideration, complicating decision-making.
7. *Communication challenges:* Communicating ProbRA results and their impact on decision and policy options can be complex. The probabilistic nature of the results may be difficult for some stakeholders to understand, posing challenges for many legal environments.
8. *Minimum data requirements:* The minimum data requirements for ProbRA are currently a topic of debate. Quantitative risk estimates are only meaningful when sufficient, statistically significant data from similar events are available and analyzed using a common criterion.
9. *Risk-benefit analysis:* Quantifying and weighing risks and benefits in a ProbRA framework can be difficult, as it requires assigning probabilities and magnitudes to both positive and negative outcomes.
10. *Diverse community perspectives:* Various communities involved in ProbRA, such as regulators, industry, and academics, may have unique sets of perspectives, historical practices, terminologies, and decision-making goals. Aligning these diverse viewpoints to achieve a harmonized approach to ProbRA can be challenging.

Addressing these challenges (Fig. 2) will require collaborative efforts among stakeholders to establish best practices, develop clear communication strategies, and build capacity for ProbRA in toxi-

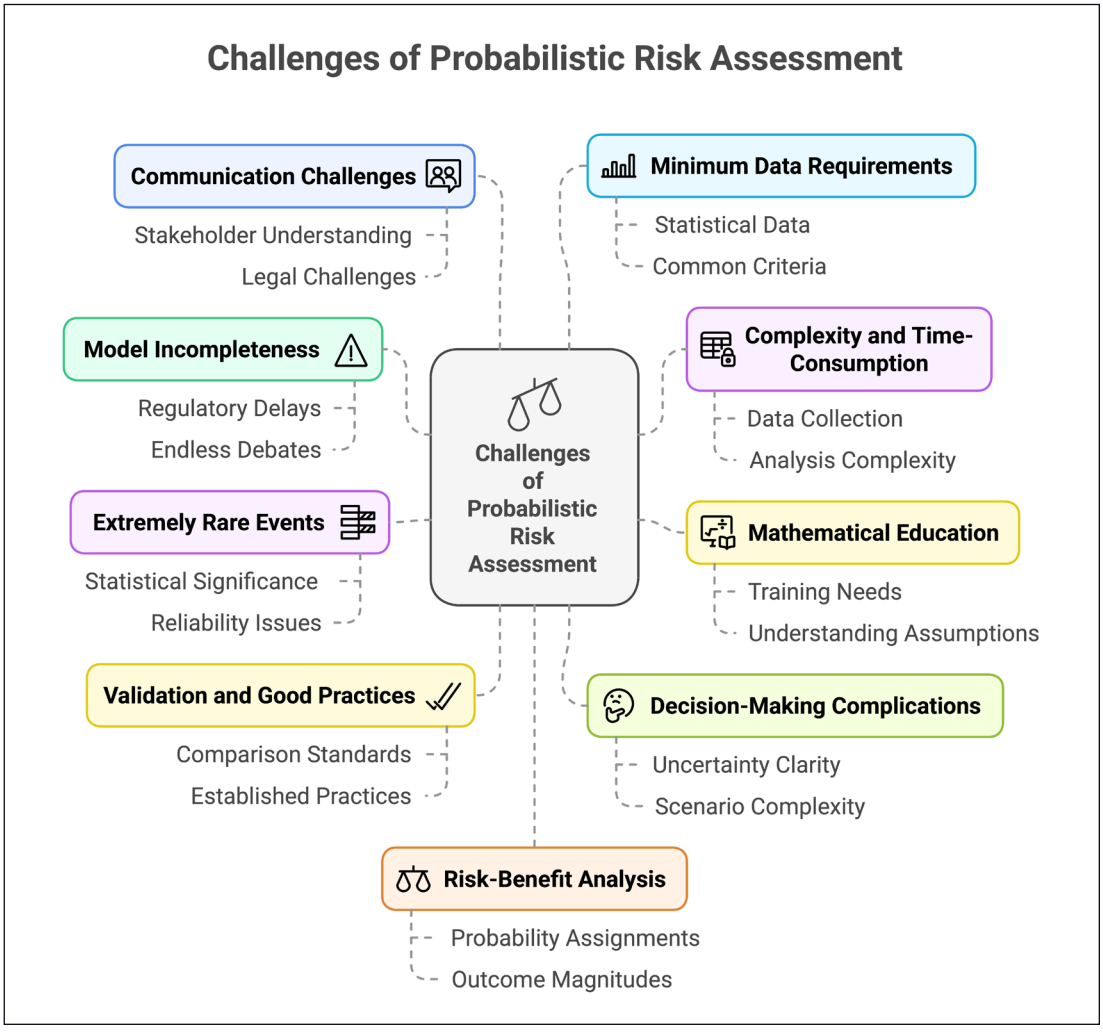

Fig. 2: Challenges of probabilistic risk assessment Designed with napkin.ai.

cology. By acknowledging and working to overcome these challenges, the toxicology community can unlock the full potential of ProbRA for improved risk assessment and decision-making.

5 Chemical risk assessment: Principles and moving towards probabilistic approaches

Chemical risk assessment plays a crucial role in ensuring food safety and protecting public health. A presentation from the European Food Safety Authority (EFSA) explored the principles of chemical risk assessment and discussed the questions that need to be addressed to move towards probabilistic approaches, with a focus on applied risk assessment in the context of food safety.

Risk assessment follows a structured, tiered approach that is designed to be fit-for-purpose. The key steps in the risk assessment process include:

1. Hazard identification: identifying the toxic effects of a chemical
2. Hazard characterization: establishing the dose-response relationship
3. Exposure assessment: determining the extent of human exposure
4. Risk characterization: integrating the information from the previous steps to estimate the risk

The weight of evidence (WoE) approach (Linkov, 2015) is used to evaluate and integrate evidence from various sources. The EFSA WoE guidance document[6] outlines a tiered approach, with Tier 0 being applicable when little occurrence data or few consumption default values are available. In such cases, *in silico* approaches can be employed. The goal is to derive a chemical-specific adjustment factor. The WoE approach involves a three-step process:

1. Assemble evidence into lines: select, extract, and list relevant evidence

[6] https://www.efsa.europa.eu/en/efsajournal/pub/4971

2. Weigh the methods and results
3. Integrate the methods and results

AOPs (Leist et al., 2017) provide a framework for understanding the toxicological effects of chemicals. Toxicokinetics, which describe the absorption, distribution, metabolism, and excretion of a chemical, are chemical-specific. A key question is where to use qualitative and quantitative approaches in the AOP framework, considering both toxicokinetics and toxicodynamics.

5.1 Moving towards probabilistic hazard and risk assessment

When moving towards probabilistic hazard and risk assessment, it is important to consider whether a "one-size-fits-all" approach is appropriate. The approach should be fit-for-purpose and tailored to the specific question being addressed. The perspective of the risk assessor is crucial in determining the appropriate approach.

For example, mutagenicity can be expressed as a binary outcome (yes/no) or as a probability. Moving to a POD can be more challenging, as historically, *in vivo* data has been used due to regulatory requirements. The POD can be based on a no-observed-adverse-effect level (NOAEL), which is a point estimate, or a benchmark dose lower confidence limit (BMDL), which involves modeling and is quantitative. Different uncertainty factors may be applied depending on the approach.

When using new approach methodologies (NAMs) based on *in silico* methods, questions arise regarding the inclusion of uncertainty factors, the use of quantitative *in vitro-in vivo* extrapolation (QIVIVE) and physiologically-based pharmacokinetic (PBPK) modeling, and the sensitivity of different endpoints (e.g., oxidative stress, transcriptomics, metabolomics, proteomics).

The group raised questions about collaboration with ECHA in structuring toxicity data with the Organisation for Economic Co-operation and Development (OECD). Joint databases and research initiatives were discussed, highlighting the importance of taking the first step and demonstrating goodwill. Releasing PBPK models for humans, farm animals, and test animals was identified as a potential starting point, given the fact that PBPK models are well suited to translate *in vitro* concentration-response relationships to an *in vivo* dose in a virtual individual or population (McNally and Loizou, 2015). It was noted that ECHA is more interested in compliance than substances of concern, and there is goodwill for NAM-based approaches for classification and prioritization. However, different applications may require different approaches. The context of the question being addressed is crucial in determining the appropriate approach, and its uncertainties must be carefully considered and addressed if possible (Loizou, 2016).

In conclusion, chemical risk assessment is a complex process that involves the integration of evidence from various sources. Moving towards probabilistic approaches requires careful consideration of the specific question being addressed, the available data, and the appropriate modeling techniques. Collaboration between regulatory agencies and the scientific community is essential to advance the field and ensure the safety of chemicals in food and other applications.

5.2 Biological perturbation as the starting point of hazard development

Biological perturbation is central to any definition of toxicology; poisons are defined by their ability to disrupt biological systems, and Paracelsus' insight was that such disruptions are not intrinsic to a chemical but are instead a function of dose. Characterizing this dose-response (in all its complexity) was central to 20th century toxicology. Now – aided by new technologies – the biological sciences have shifted from a reductionist view to a systems biology perspective. Within toxicology this has meant an appreciation for the complexity of events leading from the initial encounter between a chemical and a molecular target to perturbed cellular signaling and failure to maintain homeostasis, organ dysfunction, and ultimately systemic disease. However, despite a growing appreciation for this complexity, conceptually, we tend to line these events up as a deterministic cascade. In reality, they are each highly stochastic. Most importantly, these are networked systems with complex feedback loops and dependencies.

In the case of cancer, the mathematical modeling of inherited cancers led to the "two-hit" hypothesis and eventually an understanding of a mutation (germline or acquired) as the initiating event (Knudson, 1996), creating a connection from the molecular level to cellular disruption and disease. It is now estimated that non-hereditary cancers require four or more distinct mutation events that result in perturbation of critical cellular signaling pathways (Chernoff, 2021). While mutagenicity, therefore, is assumed to follow a linear dose-response (although this is contested (Calabrese, 2005)), most environmental mutagens do not themselves lead to cancer in the absence of other factors – genetic, environmental, and stochastic – that tip the balance within a cell from controlled to uncontrolled cell division (Tomasetti and Vogelstein, 2015). Progression to cancer is influenced by both the genetic background of an individual as well as a multiplicity of environmental factors. The Hallmarks of Cancer now include 14 broadly defined biological processes, ranging from unlocking phenotypic plasticity to immune evasion (Hanahan, 2022). Each of these complex processes represents intricate signaling pathways, and new research has opened up the possibility of cancer acting through epigenetic mechanisms alone (Parreno et al., 2024).

This has important consequences for how we categorize chemicals, which in toxicology, and particularly regulatory toxicology, has often focused on hazard alone. While hazard represents an intrinsic property of a chemical, it can in fact only be realized in a specific context of exposure to produce risk. This is especially true for endpoints that are complex, such as carcinogenicity. Yet chemicals are still classified – in ways that have significant consequences both in terms of legislation and consumer perception – as being either carcinogens or non-carcinogens. This makes as much sense as referring to guns, cars, and peanut butter as "dangerous". While calling guns dangerous is generally accepted, labeling cars as such is context-dependent and requires qualification, and referring to peanut butter as dangerous is only accurate in cases of severe individual allergies. Indeed, even though the IARC is a hazard-based assessment, the need to qualify many

potential hazards with exposure-type context has found its way into IARC's evaluations (e.g., drinking hot beverages[7], manufacturing aluminum).

Because of the precautionary principle, the use of mechanistic data, which can be prone to false positives (Trosko and Upham, 2005), and an exclusive focus on hazard versus risk, IARC classifications are at times at odds with the assessments done by other agencies, as is the case with glyphosate (Davoren and Schiestl, 2018). In its recent update, the IARC removed "Category 4 – unlikely to be carcinogenic" (which had only one chemical, caprolactam), implying that every chemical is either carcinogenic at some level, or simply not classifiable. Perhaps it is the case that for a vulnerable person under a highly specific set of circumstances, most chemicals probably have some theoretical potential to cause cancer. On the other hand, if everything causes cancer, that ceases to be a useful label – and at some point, an improbable risk is functionally equivalent to impossible. Adding to this, the cancer assay accumulates false-positive results with group size. Gaylor et al. (2005) showed that in a series of 156 chemicals tested in the National Toxicology Program, raising the number of animals per chemical from 50 to 200 would increase the number of positive chemicals from 62% to 92%. So, a non-carcinogenic chemical has apparently just not been tested enough.

Even more straight-forward endpoints, such as skin sensitization (which has one MIE – covalent binding to a protein) can be difficult to regulate with a binary classification without regard of potency. Exposure patterns (frequency and duration) in part determine potency, with considerable variation in individual genetic susceptibility determining response (Bønnelykke et al., 2015), all of which are difficult to measure and are often not reflected in regulatory regimes. The Globally Harmonized System (GHS) classifies skin sensitizers primarily into Category 1, with sub-categories 1A and 1B based on the frequency and potency of sensitization observed in humans and animals. 1A is defined as inducing skin sensitization in a substantial number of exposed individuals, yet "substantial" is never defined. The diversity and variability in this evidence make it challenging to establish clear and consistent criteria for predicting the potency of skin sensitizers. For example, Germany's occupational exposure limits marked "Sh" (skin sensitizer) and "Sah"[8] (skin and airway sensitizer) are typically considered evidence for high hazard of skin sensitization by other regulatory regimes. However, the criterion for sufficient evidence of a MAK (maximum workplace concentration) skin sensitizer – case reports of clinically relevant sensitization for multiple patients from at least two independent centers – may not align with "high hazard" and depends in part on exposure prevalence, surveillance and reporting (Committee on the Design and Evaluation of Safer Chemical Substitutions, 2014[9]). Not surprisingly, computational models tend to struggle when predicting an outcome such as skin sensitization as binary when the reality is considerably more complex – false positives or negatives are often weak sensitizers. In this sense, reformulating hazard prediction along probabilistic lines will likely improve model accuracy (Luechtefeld et al., 2015; Golden et al., 2023).

Developmental (neuro)toxicity is another example of the need to add both mechanistic grounding (EFSA, 2021) and probabilistic nuance to our understanding of chemical cause and effect: This is an area where animal tests have often failed to predict human results, and considerable variability in outcome has hampered our ability to definitively identify hazardous chemicals (Smirnova et al., 2014, 2024; Debad et al., 2025; Cöllen et al., 2025; Celardo et al., 2025). Valproic acid – used as a positive control for developmental neurotoxicity studies – shows a vast range of effects in exposed human fetuses, ranging from minor cognitive effects to spina bifida, requiring large population studies to accurately estimate the odds ratio of outcomes such as autism (Christensen et al., 2013). Understanding the connection between valproic acid and adverse outcomes has been complicated by the fact that exposures are typically concentrated in mothers with epilepsy who may have increased genetic susceptibility to some neurological outcomes as well as a different risk profile and are often taking other medications, as polytherapy is common in epilepsy (Bjørk et al., 2022). The diversity in outcomes is likely the result of many factors: genetic diversity in terms of metabolism, transport and target proteins; placental function, which can affect transfer and metabolism; and timing of exposure – with exposure in the first trimester likely causing more significant congenital defects resulting from failure of neural tube closures compared to later exposures; dose of the drug and duration are also significant factors. Several environmental exposures such as smoking, alcohol, and diet can interact with valproic acid; folate consumption may be protective (Reynolds and Green, 2020). Finally, fetal gender also impacts outcomes, with males potentially more vulnerable than females. These effects have been challenging to ascertain, even though exposures tend to be consistent, relatively high, and well documented via medical records. This is in marked contrast to most exposures to environmental chemicals – there remains considerable controversy about other potential developmental neurotoxicants, such as flame retardants (Blum et al., 2019).

Therefore, a linear AOP – a series of dominos, one knocking down the other – must be replaced with something more akin to a Waddington landscape (Fig. 3). At a cellular level, this can be conceptualized as a steady-state homeostatic cell having the potential to be perturbed by an external agent into an altered (pathological) state. More formally, if we reconsider normal and tumor cell types as attractors, then chemical exposures shift the height of hills between two valleys, or the barrier heights between the two attractors. Unlike the developmental landscape, the possibility of returning to steady-state homeostasis can persist. The probability that a chemical will alter cell fate by causing an escape from the basin of attraction serves as a quantitative measure of hazard at the cellular level. This measure is analogous to potency and respects

[7] https://www.iarc.who.int/wp-content/uploads/2018/07/pr244_E.pdf

[8] https://safety-work.org/fileadmin/safety-work/articles/Grenzwerte_fuer_Gefahrstoffe_am_Arbeitsplatz/Grenzwerte_fuer_Gefahrstoffe_am_Arbeitsplatz.pdf

[9] https://www.ncbi.nlm.nih.gov/books/NBK253961/

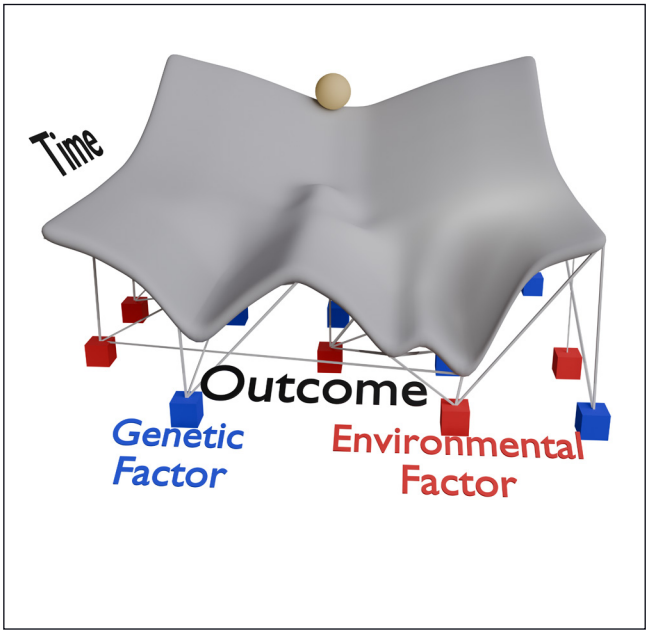

Fig. 3: Waddington landscape of tumor development
The Waddington landscape of tumor development (Aranda-Anzaldo and Dent, 2018) refers to a metaphorical representation, based on the concept of an "epigenetic landscape" proposed by Conrad Waddington, which visualizes the progression of a normal cell into a cancerous cell as a ball rolling down a complex terrain, where different valleys represent distinct cell fates, with the "downhill" direction signifying increasing malignancy, highlighting how genetic and environmental factors can influence a cell's trajectory towards becoming cancerous through various bifurcations and attractor states along the way.
A pluripotent stem cell can differentiate into either a normal cell or a cancer cell, influenced by the combined effects of genes (blue) and environmental agents (red). While many germline and somatic mutations that significantly lower the energy barrier between normal and cancer cells have been cataloged, our understanding of weaker gene and environmental interactions remains limited. A molecular initiating event (MIE) can push the cell away from its normal attractor state to an unstable intermediary. However, the progression to the cancer attractor state also depends on genetic predisposition.

the fundamental stochasticity of the process, the complexity, and the considerable genetic individual variability in the landscape.

Clearly, dosing small mammals at high levels with a toxicant gives us little, if any, information to fill out this model. At best, it gives an estimate of the toxicokinetics, but little information about the toxicodynamics at the cellular level, often showing only organ pathology but revealing little about the MIE or subsequent perturbed signaling. It often leaves considerable uncertainty about species differences, as well as the susceptibility of vulnerable populations. This has been previously dealt with by applying uncertainty factors of 10 for each source of uncertainty. Yet the continued reliance on large uncertainty factors – which dates back to 1954 (Dankovic et al., 2015) – can be seen as an implicit admission that we cannot, with any exactitude, predict chemical impacts in humans.

What types of data would give us more realistic metrics? In other words, how do we establish what is the "tipping point" – the critical threshold at which a small perturbation within a cell can cause a significant shift from one state to another, and eventually pathology?

5.3 The adverse outcome pathway concept to address perturbation of biology

The AOP concept was suggested out of an OECD ecotoxicology working group (Ankley et al., 2010) but was quickly expanded to include human hazards (Leist et al., 2017). Parallel efforts toward pathways of toxicity (PoTs) (Hartung and McBride, 2011; Kleensang et al., 2014) were developed within the Human Toxome Project (Bouhifd et al., 2014, 2015).

5.3.1 From molecular initiating events to population

1. Molecular initiating event (MIE)

To begin with, every chemical initiates a toxic response in a limited number of MIEs, e.g., inhibiting an enzyme, covalently binding to a macromolecule, binding to a receptor, or affecting membrane permeability. In theory, this interaction is determined by the chemical structure – electronegativity determines the likelihood of binding to DNA or a protein; the 3D shape determines receptor binding. It follows that we should be able to predict MIEs from structure alone – and in many instances we can. Many existing models based on (quantitative) structure-based activity relationships ((Q)SARs) are approaching an accuracy level similar to animal models, and as larger datasets that explore the chemical space become available, accuracy is expected to increase (Luechtefeld et al., 2018a; Golden et al., 2023). Similarly, molecular docking can predict receptor binding, and AI will likely improve not just our accuracy but will increase the scale of available data in the same manner that AlphaFold has increased our understanding of the known protein universe (Jumper et al., 2021). A decade ago, performing bulk target receptor binding assessment of extensive chemical libraries, as made possible by the AI-guided DeepDock integration of the publicly available Vina AutoDock software (Trott and Olson, 2010), would have been far too computationally intensive. In addition, data on MIEs are often the easiest to obtain at a scale that makes "big data" approaches possible.

The likelihood of a chemical binding to its target is intrinsically stochastic and depends on the concentration, affinity, and the presence of competing molecules. Each of these factors is influenced by considerable natural variability in biological systems as well as co-exposures. In lieu of a dose-response with a fixed threshold, the dose-response curve should represent the likelihood of a MIE occurring at different exposure levels.

2. Key events

Key events (KEs) are more challenging to predict, model and measure. While MIEs are limited to a molecular scale with direct interactions between chemicals and biological targets, KEs propagate through various biological levels and can involve interactions across multiple scales, from molecular changes to cellular

responses, tissue effects, and ultimately organ pathology. The processes are more complex and can involve both feedback loops and homeostatic mechanisms. This multiscale nature requires integrating data from various levels of biological organization, increasing the difficulty of prediction. Biological variability in MIEs is mostly limited to toxicokinetics and is less influenced by individual variability and biological context since they depend primarily on chemical structure and target interaction; KEs are highly variable and context-dependent, influenced by individual differences in genetics, epigenetics, health status, and other environmental exposures.

This challenge, however, is also an opportunity – KEs are where our expanded repertoire of *in vitro* assays has the most potential. The past decade has seen omics technologies – particularly transcriptomics – deployed to map genetic regulatory networks and understand cellular physiology as well as signaling pathway perturbations (Kleensang et al., 2014; Maertens et al., 2015). Extending this to capture other biological layers such as proteomics, metabolomics, and methylation is only a matter of improving analytical methodologies and algorithms (Maertens et al., 2017). In addition, we have also seen the development of increasingly sophisticated read-outs of cellular phenotypes for oxidative stress, stemness, and pathology, with technologies ranging from traditional dye-based methodologies to sophisticated AI-based readouts of morphology (Chandrasekaran et al., 2024). Most importantly, we have seen significant improvement in our ability to model normal human physiology *in vitro*. Previous tissue culture assays were limited to cancer-derived cell lines that were genomically unstable, prone to artifacts, and lacked reproducibility (Kleensang et al., 2016; Pamies and Hartung, 2017; Tran et al., 2021). Tissue cultures are often cultivated under altered oxygen levels compared to normal tissue (Place et al., 2017), can be prone to artifacts from the glass or plastic surface they are grown on (Dehaan, 1971), are grown in highly artificial medium, and are not exposed to the mechanical forces they would experience *in situ* (Vertrees et al., 2009). However, with the development of improved tissue culture methodologies (ranging from providing physiological oxygenation levels to cultivation in 3D), we can better approximate human physiology. Finally, with MPS that can closely approximate normal organ function and the use of induced pluripotent stem cells (iPSCs), which can be derived from a target population, we can now capture population-level diversity and explore gene-environment interactions in the laboratory instead of through epidemiology (Suciu et al., 2023).

In silico toxicology can provide important assistance in understanding how MIEs propagate through biological networks to produce adverse outcomes. This approach moves beyond traditional linear AOPs to capture the inherent complexity in the form of networks and the stochastic nature of biological responses. Modern multi-scale modeling frameworks integrate processes across different biological scales – from molecular interactions (Montagud et al., 2022) to whole-body pharmacokinetics (Sluka et al., 2016; Letort et al., 2019) – providing a more comprehensive understanding of toxicological processes. By incorporating feedback loops, compensatory mechanisms, individual-specific variabilities, and temporal dynamics, these models can better represent how initial perturbations may lead to varying outcomes across populations with different genetic backgrounds.

A particularly powerful implementation of this approach utilizes probabilistic Boolean networks that can be derived from molecular physiological or disease maps (Grabowska et al., 2021; Hemedan et al., 2023, 2024). These maps capture the complexity of biological systems and can serve as a direct basis for dynamic models leveraging systems biology tools (Niarakis et al., 2023). Integration of omics data can bring node (i.e., genes, proteins, phenotypes, KEs) variability into model parameters, allowing specifying personalized models that can be assembled into a virtual population for stochastic hazard prediction (Ross et al., 2018; Montagud et al., 2022; Hemedan et al., 2024). The resulting models enable three key types of analysis for hazard characterization:

- *Trajectory analysis:* Predicts the probability of activation for specific endpoints, represented as biological process nodes in a network. This provides quantitative predictions of adverse outcome likelihood after the system reaches a stable state.
- *Sensitivity analysis:* Identifies critical biological entities that significantly influence progression toward toxicological phenotypes, helping elucidate key mechanistic drivers of adverse outcomes.
- *Attractor mapping:* Maps the long-term behavior of the system under different initial conditions (i.e., perturbed, not perturbed, multiple targets perturbed at different levels), revealing possible steady states and the mechanistic paths leading to them.

This framework is particularly valuable for chronic exposures, where the interaction between exposure dosage, duration, and biological adaptation becomes critical. For example, Boolean modeling has been applied to study Parkinson's disease, simulating the long-term effects of protein aggregation and cell dysfunction characteristic of chronic exposure to disease-causing factors (Hemedan et al., 2023).

Homeostatic adaptation in liver toxicity illustrates how cells initially mount protective responses against harmful exposures before these defenses become overwhelmed. This is exemplified in acetaminophen toxicity, where Phase I and II metabolic pathways and protective mechanisms like glutathione conjugation can become overwhelmed during overdose scenarios, leading to cellular damage (Sluka et al., 2016). Multi-scale modeling frameworks enable integration of ordinary differential equations (ODEs) describing chemical-target interactions, which can be combined with Boolean network models, allowing simulation of both acute and chronic – including repeated-dose – exposure scenarios (Letort et al., 2019) and simulate outcomes for different biological organization levels, from molecular dynamics to individual full body responses (Sluka et al., 2016).

3. Population level

One of the most significant success stories has been the development of precision medicine in oncology, made possible by the widespread sequencing of tumors. This has led not only to improved therapies that can target treatments to the genomic profile of both the patient and the tumor – but has expanded our understanding of the complexity of cancer in ways that can be leveraged for environmental health. Distinct mutational signatures of environmental

agents such as benzene and UV radiation can be used as biomarkers of exposure and effect, while mutational signatures of oncogenes provide a glimpse into the MIE or genetic susceptibility (Alexandrov et al., 2020). For example, a recent study demonstrated that skin cancer in a European-derived population reflected UV-radiation mutational signatures, while in an Asian population it reflected mutations in the NOTCH signaling pathway (King et al., 2023). AI-enabled reconstruction of germline mutation status from sequenced tumors (Sun et al., 2018) has clarified the effects of heterozygosity for the MUTYH gene, which were previously inconclusive due to small sample sizes and the high cost of genotyping in epidemiology studies. This method revealed that while heterozygosity for the MUTYH gene does not appear to increase susceptibility to hormonally-driven cancers, it does elevate the incidence of other cancer subtypes. Consequently, researchers can now concentrate on more likely gene-environment interactions (Paller et al., 2024).

In some respects, making better use of human data will bring toxicology back to its origins. Most human carcinogens were originally discovered based on the observation of higher rates of cancers in occupationally exposed individuals (Weindling, 1985). The Ames test (and other *in vitro* tests of mutagenicity and genotoxicity) shifted focus to the MIE but also demonstrated that mutagens were in fact very common (Ames and Gold, 1990). The usefulness of the rodent model for predicting cancer remains up for debate (Gottmann et al., 2001; Osimitz et al., 2013; Basketter et al., 2012). What is beyond dispute is the vast improvement large-scale human data sets have brought to our understanding of the causes, complexity, and treatment of cancer (Tomczak et al., 2015).

Cancer, uniquely, has the advantage of allowing for an easy molecular read-out of pathological tissue – a tumor can now be sequenced as a routine part of care, and more in-depth datasets with mutational profile, transcriptomics, and methylation, are readily available for AI-approaches. Translating this to diseases with less accessible pathology – for example, neurodevelopmental and neurodegenerative diseases – will require a different approach. Yet here too there are lessons to be learned from recent successes – the mining of electronic health records for exposures that can be connected to disease has been able to definitively establish both valproic acid as a causal agent for autism (Jentink et al., 2010) and anticholinergic drugs for Alzheimer's disease (Coupland et al., 2019) – in the latter case, grouping drug exposures by MIE and duration of exposure was crucial for establishing the statistical significance of long-term effects. A more expansive catalog of MIEs and improved biomarkers for KEs will therefore improve our ability to connect cause and effect in large-scale population data sets of exposures and outcomes.

6 Systems toxicology and its relation to identifying the probability of hazard from perturbation of biology

Systems toxicology, an extension of systems biology, integrates high-throughput omics data (genomics, transcriptomics, proteomics, and metabolomics) with computational modeling to understand the complexity of biological responses to chemical exposures (Hartung et al., 2017). This holistic approach aims to elucidate the intricate networks of molecular interactions and pathways perturbed by toxicants, providing a more comprehensive understanding of how these perturbations lead to adverse health outcomes.

Systems toxicology leverages advanced technologies such as high-content screening and high-throughput assays to generate vast amounts of data. These data are used to create detailed biological profiles, known as signatures of toxicity (SoT). These signatures capture the immediate and downstream effects of chemical exposure on biological systems, allowing researchers to identify pathways of toxicity (PoT, Kleensang et al., 2014) that lead to adverse health effects. Mapping pathway perturbations involves integrating data from various omics technologies. For example, transcriptomics can reveal changes in gene expression levels, while proteomics can identify alterations in protein abundance and modifications. Metabolomics provides insights into changes in metabolic pathways. Combining these data can identify critical nodes and interactions within biological networks that are perturbed by toxicants.

Systems toxicology has the potential to revolutionize regulatory toxicology by providing more accurate and predictive models of chemical safety. Key applications include:

- *Chemical screening:* High-throughput screening of chemicals to identify potential hazards based on their signatures of toxicity.
- *Mechanistic insights:* Elucidating the mechanisms underlying chemical-induced toxicity, which can inform the development of safer chemicals and targeted interventions.
- *Risk assessment:* Improving the accuracy and reliability of risk assessments by incorporating probabilistic models and pathway-based data.

The future of systems toxicology lies in the integration of diverse data sources and the development of sophisticated computational models that can simulate complex biological responses. Advances in machine learning and AI (data-driven, black box approaches) will play a crucial role in this endeavor, enabling the prediction of adverse outcomes with greater precision. On the other end of the *in silico* spectrum, mechanistic models (knowledge-driven, white box approaches) are already being used to predict adverse outcomes from a molecular perspective. In between, a combination of all "shades" of modeling (grey box models) will allow employing the optimal combination of data and knowledge and will potentiate our ability to predict adverse effects and personalize risk assessment from a population to the individual level.

In conclusion, systems toxicology represents a paradigm shift in the way we assess chemical hazards. By focusing on the perturbation of biological pathways and adopting probabilistic approaches, it offers a more comprehensive and accurate assessment of risk. This approach not only enhances our understanding of toxicological mechanisms but also provides a robust framework for the development of safer chemicals and more effective regulatory policies.

7 The challenge of modeling biological systems

Modeling biological systems is a complex and multifaceted task that requires addressing several key challenges. These challenges arise from the inherent complexity of biological networks, the

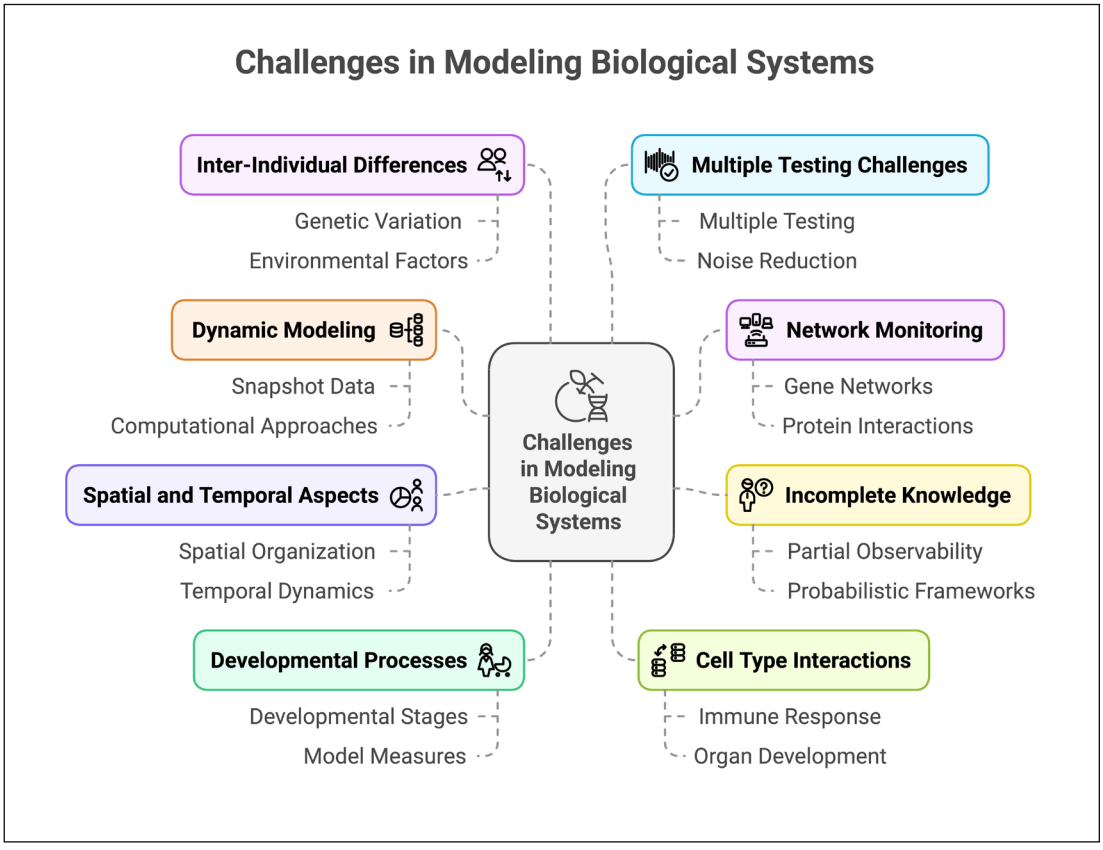

Fig. 4: Challenges in modeling biological systems Designed with napkin.ai.

limitations of current experimental techniques, and the dynamic nature of living systems. This chapter explores these challenges and discusses potential strategies for overcoming them.

7.1 The challenges

A number of problems have to be tackled (Fig. 4):

a) Dynamic modeling with snapshot data

One of the primary challenges in modeling biological systems is the need for dynamic modeling. Biological processes are inherently dynamic, with complex interactions and feedback loops that evolve over time. However, experimental techniques often provide only snapshots of the system at specific time points. To construct accurate dynamic models, researchers must combine these snapshots with existing knowledge of reaction kinetics and use computational approaches to infer the underlying dynamics. This requires the development of sophisticated algorithms and the integration of data from multiple sources, such as time-series experiments, literature mining, and expert knowledge.

b) Compatibility of network monitoring technologies

Biological systems are composed of multiple interacting networks, such as gene regulatory networks, protein-protein interaction networks, and metabolic pathways. Each of these networks has its own set of technologies for monitoring and data acquisition, such as RNA sequencing for gene expression, mass spectrometry for protein abundance, and metabolomics for metabolite profiles. A significant challenge lies in the fact that these technologies are not always compatible for measuring the same sample at the same time. This incompatibility can lead to difficulties in data integration and the need for separate experiments, potentially introducing unwanted variability and limiting the ability to capture the true dynamics of the system.

c) Incomplete knowledge and measurability of systems

Many biological systems are not completely known or measurable, with only a subset of network components being accessible for monitoring. This partial observability poses challenges for modeling, as the behavior of the system may be influenced by unmeasured or unknown factors. Researchers must develop strategies to infer the states of hidden variables and account for the uncertainty introduced by incomplete knowledge. This can involve the use of probabilistic modeling frameworks, such as Bayesian networks or hidden Markov models, which allow for the incorporation of prior knowledge and the estimation of latent variables.

d) Spatial and temporal aspects of biological systems

Biological systems often exhibit spatial and temporal organization that is critical for their function. Spatial aspects, such as the compartmentalization of cellular processes within organelles or the localization of molecules in specific tissues, can have significant impacts on the behavior of the system. Similarly, temporal aspects, such as the sequence and timing of events in signaling cascades or the circadian rhythms of gene expression, play crucial

roles in regulating biological processes. Incorporating these spatial and temporal features into models requires the development of specialized computational frameworks, such as partial differential equation models or agent-based models, which can capture the complex dynamics and interactions within the system.

e) Developmental aspects and model measures of interest
Biological systems undergo developmental processes, such as establishment, maturation, differentiation, and aging, which can have profound effects on their behavior and response to perturbations. These developmental aspects must be considered when constructing models, as they can influence the parameters and structure of the network over time. Additionally, the measures of interest in a model may change depending on the developmental stage or the specific question being addressed. For example, a model of embryonic development may focus on the spatial patterning of gene expression, while a model of aging may prioritize the accumulation of cellular damage and the decline of functional capacity.

f) Interactions of different cell types and tissues
Many physiological processes involve the interaction of multiple cell types and tissues, adding layers of complexity to the modeling process. For example, the immune response involves the coordinated action of various immune cell types, each with its own set of signaling pathways and regulatory mechanisms. Similarly, the development and function of organs depend on the interplay between different cell types, such as epithelial cells, stromal cells, and vascular cells. Modeling these interactions requires the integration of data from multiple sources and the development of multi-scale models that can capture the emergent properties arising from the interplay between different levels of biological organization.

g) Inter-individual differences and data acquisition
Inter-individual differences, arising from genetic variation, environmental factors, and stochastic processes, can have significant impacts on the behavior of biological systems. These differences can affect the parameters and structure of the network, leading to variability in the response to perturbations or interventions. Accounting for this variability in models requires the acquisition of data from diverse populations and the development of personalized or stratified modeling approaches. However, obtaining sufficient data to capture inter-individual differences can be challenging, as it may require large sample sizes and the integration of data from multiple cohorts or studies.

h) Multiple testing, over-fitting, and noise/signal ratios
The complexity of biological systems and the high-dimensional nature of the data generated by modern experimental techniques pose challenges related to multiple testing, over-fitting, and noise/signal ratios. As the numbers of parameters and conditions in a model increase, the risk of false positives due to multiple testing and over-fitting of the data also rises. Additionally, biological measurements are often subject to technical and biological noise, which can obscure the true signal and make it difficult to identify the underlying mechanisms. Addressing these challenges requires the use of rigorous statistical methods, such as false discovery rate control and cross-validation, as well as the development of robust and interpretable modeling frameworks that can distinguish between true biological signals and spurious associations. In addition, combining data-driven approaches with mechanistic ones through advanced *in silico* modeling can help to amplify the signal.

7.2 Strategies for overcoming the challenges

To address the challenges of modeling biological systems, researchers must adopt a multidisciplinary approach that combines experimental, computational, and theoretical techniques. Some strategies for overcoming these challenges include:

1. *Integration of multi-omics data:* Combining data from different levels of biological organization, such as genomics, transcriptomics, proteomics, and metabolomics, can provide a more comprehensive view of the system and help to overcome the limitations of individual technologies.
2. *Development of advanced computational methods:* Novel computational approaches, such as machine learning, deep learning, and network inference algorithms, can help to uncover hidden patterns and relationships in the data and enable the construction of more accurate and predictive models.
3. *Incorporation of prior knowledge:* Integrating existing knowledge from the literature, databases, and expert opinion can help to constrain the model space and improve the interpretability and robustness of the results. In addition, efforts to curate species-specific pathways can help in understanding interspecies variances and addressing these differences in modeling, making outcomes more relevant to humans and moving away from benchmarking NAM predictions against classical animal studies.
4. *Multi-scale and hybrid modeling:* Combining different modeling frameworks, such as ordinary differential equations, partial differential equations, agent-based models, and logic-based models (e.g., Boolean networks), can enable the capture of both the spatial and temporal aspects of biological systems and the interactions between different levels of organization.
5. *Personalized and stratified modeling:* Developing models that account for interindividual differences and stratify populations based on relevant factors, such as genetic background or environmental exposures, can improve the accuracy and applicability of the results.
6. *Rigorous model validation and uncertainty quantification:* Employing robust statistical methods for model validation, such as cross-validation and bootstrapping, and quantifying the uncertainty in the model predictions can help to ensure the reliability and reproducibility of the results. Here again, the discussion on species relevance must be in the spotlight. Human-relevant models should be validated against human-relevant data, or at least interspecies variabilities should be accounted for.

In conclusion, modeling biological systems is a challenging task that requires addressing the inherent complexity of living systems, the limitations of current experimental techniques, and the dynamic nature of biological processes (Fig. 5). By integrating multi-omics data, developing advanced computational methods,

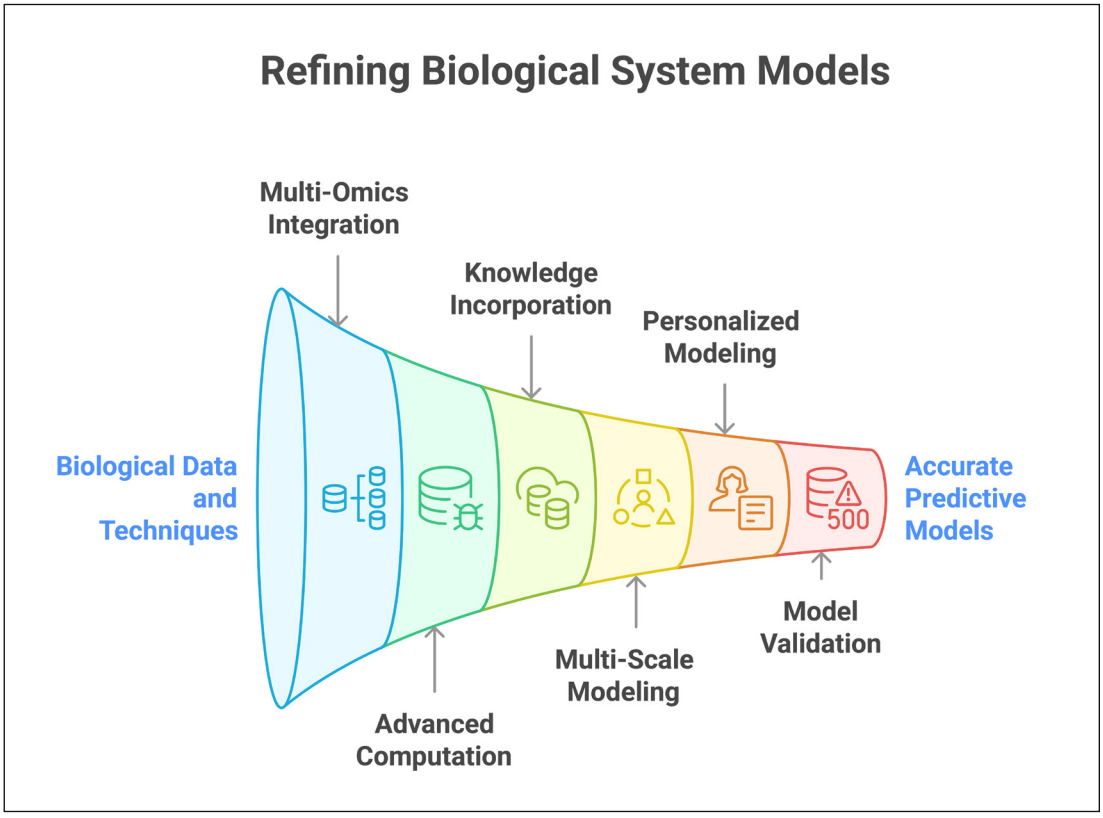

Fig. 5: Refining biological systems models Designed with napkin.ai.

incorporating prior knowledge, and employing rigorous validation and uncertainty quantification techniques, researchers can overcome these challenges and construct more accurate, predictive, and interpretable models of biological systems. These models can provide valuable insights into the mechanisms underlying health and disease and guide the development of new therapies and interventions. As the field of systems biology continues to evolve, the integration of experimental, computational, and theoretical approaches will be essential for unraveling the complexity of biological systems and advancing our understanding of life at the molecular, cellular, and organismal levels.

7.3 Virtual experiments in systems approaches to challenge our understanding of causality

Systems biology and computational modeling are revolutionizing our understanding of biological processes and how perturbations can lead to adverse outcomes. By integrating experimental data into computational models, we can gain insights into the complex interactions and emergent properties of biological systems that are difficult to discern through traditional reductionist approaches (Smirnova et al., 2018). These models allow us to perform virtual experiments to test hypotheses and generate predictions that can guide further experimental work. However, this approach also challenges our traditional notions of causality and requires a shift in thinking about how we establish causal relationships in complex systems.

The promise of systems toxicology lies in its ability to capture the complexity of biological systems and provide a more human-relevant approach to safety assessment. By building computational models based on human cell cultures, organoids, and microphysiological systems, we can better recapitulate human physiology and predict toxicological responses. These models integrate data across multiple levels of biological organization, from MIEs to cellular responses, tissue-level effects, and ultimately organ and organism-level outcomes. By considering the full spectrum of biological interactions and feedback loops, systems models can provide a more comprehensive and mechanistic understanding of how chemical exposures lead to adverse effects.

However, the complexity of these models also presents challenges for establishing causality. In traditional toxicology, causality is often inferred from observed associations between exposures and outcomes, with the assumption that intervening variables have been adequately controlled. But in complex systems with numerous interacting components, it can be difficult to isolate specific causal relationships and rule out alternative explanations. Systems models may reveal unexpected emergent properties or non-linear dose-response relationships that complicate causal inference.

To address these challenges, systems toxicology relies on an iterative process of model building, virtual experimentation, and experimental validation (Smirnova et al., 2018). Computational models are constructed based on existing knowledge and experimental data and then used to generate predictions about the sys-

tem's behavior under different conditions. These predictions can then be tested experimentally, with the results used to refine and improve the model. By cycling between virtual and physical experiments, researchers can progressively build confidence in the model's ability to capture causal relationships and generate reliable predictions.

This process also requires a more nuanced view of causality that considers the probabilistic and context-dependent nature of biological systems. Rather than seeking to establish deterministic causal chains, systems toxicology aims to identify the key drivers and modulators of adverse outcomes and characterize their interactions. This may involve considering multiple causal pathways, feedback loops, and contextual factors that influence the likelihood and severity of effects. Techniques such as sensitivity analysis and uncertainty quantification can help to evaluate the robustness of causal inferences and identify areas where additional data or model refinements are needed.

Ultimately, the goal of systems toxicology is to develop a predictive understanding of how chemical exposures impact human health, based on a mechanistic understanding of the underlying biological processes. By integrating virtual and physical experiments in an iterative fashion, this approach can generate more reliable and human-relevant predictions of toxicological risk. However, it also requires a willingness to embrace the complexity of biological systems and adopt new ways of thinking about causality and evidence. As the field of systems toxicology continues to evolve, it will be important to develop rigorous methodologies for model validation, uncertainty quantification, and causal inference that can build confidence in this approach and support its application in regulatory decision-making. A strong starting point is learning from the American Society of Mechanical Engineering (ASME)[10] Verification, Validation, and Uncertainty Quantification VVUQ protocols, some of which already cover computational modeling of medical devices (V&V-40), machine learning models (V&V-70), and pharmaceutical products (V&V-80). Currently, an ISO-IEC working group is preparing a harmonized version of the V&V-40. In addition, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is currently working on model-informed drug development, revising the inclusion of various modeling approaches and their credibility assessment.

7.4 Advancing probabilistic risk assessment in toxicology

The field of toxicology is undergoing a paradigm shift towards the integration of advanced methodologies, such as AI, *in silico* trials, and digital twins, to improve the accuracy and relevance of risk assessment. This chapter explores how traditional methods inform AI, the synergies between these approaches, and the potential for leveraging emerging technologies to advance ProbRA in toxicology.

Synergy between traditional methods and AI: Traditional toxicological methods, such as *in vivo* animal studies and *in vitro* assays, provide valuable data and mechanistic insights that can inform the development of AI models (Hartung, 2023b,c; Kleinstreuer and Hartung, 2024). By leveraging the knowledge gained from these methods, AI can help identify patterns, predict outcomes, and extrapolate findings to human populations. However, the full potential of this synergy has not yet been fully exploited, presenting opportunities for further advancement. A major challenge is the validation of these AI-based approaches (Hartung and Kleinstreuer, 2025).

In silico *trials and digital twins: In silico* trials, which involve creating virtual cohorts of digital twins based on clinical data, are increasingly being used in the pharmaceutical industry for drug development and device testing (Craig et al., 2023; Bordukova et al., 2024). Adapting this approach to toxicology could enable the prediction of population-level effects from subcellular-level perturbations caused by chemicals. By integrating knowledge of how chemicals affect gene expression and molecular pathways, digital twins can help translate these findings to clinical outcomes. Models in some areas, such as cardiology and cardiotoxicity, are already outperforming traditional animal models in identifying cardiotoxic mechanisms (Passini et al., 2017). For example, *in silico* drug trials using an experimentally-calibrated model of human ventricular myocyte electrophysiology and Ca^{2+} cycling (O'Hara et al., 2011) have demonstrated higher accuracy than animal models in predicting clinical pro-arrhythmic cardiotoxicity and are used by pharmaceutical industry (Passini et al., 2017). The use of digital twins represents a shift from traditional toxicology to a more probabilistic and human-relevant approach. Large sets of clinical data can be used to create virtual patient populations, allowing for the extrapolation of *in vitro* findings to human populations through modeling and simulation informed by machine learning.

Regulatory acceptance: Several studies demonstrate the potential of *in silico* methods in toxicology (O'Hara et al., 2011; Passini et al., 2017). For example, both the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) acknowledge the value of *in silico* methods, especially PBPK models, in drug development and evaluation, with digital evidence from these models being accepted in regulatory submissions (Musuamba et al., 2021). The number of submissions to the FDA containing PBPK-derived evidence increased more than 50% from 2017 to 2019 (Grimstein et al., 2019). This highlights the growing interest from industry and the increasing acceptance by regulators of these emerging methodologies.

Addressing human variability: One of the key advantages of using digital twins in toxicology is the ability to account for human variability and personalized factors. Unlike traditional animal studies, which often use genetically homogeneous rat or mouse strains, digital twins can incorporate individual differences in genetics, metabolism, and susceptibility to chemical exposures. This allows for a more comprehensive assessment of the probability of hazard at the population level.

Improving hazard assessment with biological data: Incorporating biological data into ProbRA can help distinguish between significant effects and relevant effects, as well as identify com-

[10] https://www.asme.org/codes-standards/publications-information/verification-validation-uncertainty

pensatory mechanisms and adaptive responses. By modeling the perturbation of biological systems and measuring omics data, we can gain insights into the differences in individual responses and the intrinsic factors that influence the manifestation of hazard. The concept of AOPs provides a framework for understanding the contribution of KEs to the overall network of biological responses. However, there are still missing models that bridge the gap between cellular processes and higher levels of biological organization (e.g., tissues and organs). Biologically informed models, such as those developed for drug-induced QT prolongation, demonstrate how the integration of multiple pathways can help mitigate risk and avoid overly conservative approaches. These include:

1. *Quantifying biology and setting thresholds:* To effectively incorporate biological data into ProbRA, it is necessary to quantify the extent of perturbation and determine the thresholds at which adaptive responses are overwhelmed and adverse effects emerge. This requires the identification of patterns and the minimum number of molecular drivers that characterize the tipping point between adaptation and adversity.
2. *Test cases and proof of principle:* Several test cases can be used to demonstrate the feasibility and value of ProbRA approaches in toxicology. For example, transcriptomics data from drug-induced liver injury (DILI) studies have been used to develop predictors of outcome, combining gene signatures and toxicogenomics with chemical structure information. Similarly, short-term animal studies with metabolomics data have been used to group chemicals and identify modes of action (MOAs), enabling the assessment of new substances based on their expected patterns.
3. *Integration of MPS, clinical data, and* in silico *models:* The integration of MPS, clinical data, and *in silico* models represents a promising approach for advancing ProbRA in toxicology. MPS can provide relevant biological readouts, such as transcriptomics or morphological changes, which can be used to parameterize digital twin models. The combination of MPS and *in silico* models allows for the linking of cellular-level perturbations to population-level disease outcomes.

In conclusion, the integration of traditional toxicological methods, advanced *in vitro* systems, and emerging *in silico* approaches, such as digital twins and AI-informed modeling, holds great promise for advancing ProbRA in toxicology. By leveraging the synergies between these methods, incorporating human variability, and quantifying the tipping points between adaptation and adversity, we can develop more accurate, relevant, and protective approaches to assess chemical hazards. Continued research, validation, and regulatory acceptance of these methodologies will be crucial for their successful implementation in toxicology and risk assessment.

7.5 Causation in complex networks facilitated by AI

In the context of perturbation of biology leading to hazard, understanding causation in complex networks is crucial for identifying the underlying mechanisms and predicting the potential adverse effects of exposures or interventions. Complex biological systems, such as gene regulatory networks, metabolic pathways, and signaling cascades, are characterized by intricate patterns of interactions and dependencies that can obscure the true causal relationships between variables.

Key tools and methods, such as causal graphs, causal inference, Granger causality, transfer entropy, complex network analysis, and dynamical systems theory, provide a framework for unraveling the causal relationships within these complex networks. These approaches enable researchers to visualize causality, quantify information flow, identify structural patterns, and capture network dynamics, leading to a deeper understanding of how perturbations propagate through biological systems and give rise to adverse outcomes.

AI has emerged as a transformative force in the study of causation in complex networks, offering powerful tools for analyzing the vast amounts of data generated by modern experimental techniques. AI techniques, such as machine learning, data mining, and deep learning, can uncover hidden patterns and correlations that may indicate causal relationships between different nodes in the network. Predictive modeling, causal inference, simulation and optimization, network reconstruction, and anomaly detection are among the key applications of AI in this context.

By leveraging AI approaches, researchers can identify the MIEs and KEs that link exposures to adverse outcomes, providing mechanistic insights into the perturbation of biological systems. AI-powered analysis of omics data, such as transcriptomics, proteomics, and metabolomics, can reveal the complex cascades of molecular changes that occur in response to chemical exposures or other perturbations. These insights can inform the development of AOPs and support the prediction of hazard based on an understanding of the underlying biological mechanisms (de Vries et al., 2021).

Moreover, AI techniques can help characterize the complex interactions between exposures and genetic factors that determine an individual's susceptibility to adverse effects. By studying the perturbation of biological networks in genetically diverse populations, researchers can better understand gene-environment interactions and identify vulnerable subgroups who may be at higher risk of exposure-related diseases.

In conclusion, the integration of key tools and methods for understanding causation in complex networks, coupled with the power of AI, holds immense promise for advancing our understanding of how perturbations in biological systems lead to hazard. By unraveling the intricate web of cause and effect within these networks, researchers can gain mechanistic insights, develop predictive models, and inform risk assessment and decision-making in the context of chemical safety and public health. However, careful consideration must be given to the limitations and biases of these approaches, and collaboration between domain experts and AI specialists is essential to ensure the validity and reliability of the insights generated.

8 Challenges to the new paradigm

This new paradigm will come with challenges – there is a need for better analytical methodologies, assembling and curating large-scale data sets – in particular using text-mining, not just to organ-

ize the existing literature-base but to include both regulatory data and gray literature, such as safety data sheets, to overcome our publication bias. It will certainly require the use of AI to overcome our limited understanding of biological interactions and shift to a systems biology perspective (Krewski et al., 2010; Kleensang et al., 2014). But it will also require embedding an understanding of ProbRA in how we talk about hazard. The distinction between mutagenicity and carcinogen potency is well-understood, yet the failure to fully implement this in regulatory regimes has led to such absurdities as the State of California insisting that coffee required a warning label about its potential for carcinogenicity, which was only removed after a lengthy legal battle[7] (Flynn, 2019). No doubt, large-scale data sets and molecular docking will reveal a wealth of substances, both natural and artificial, that have some level of estrogen receptor binding; we cannot declare all of them endocrine disruptors (Cozzini et al., 2022).

Similarly, omics data is inherently upstream of phenotypic manifestations of hazard. What portion of differentially expressed genes, translated proteins, generated metabolites, or altered subcellular functionality should be deemed the minimum acceptable evidence of an adverse outcome? How do we benchmark these when used as biomarkers of effect? Moreover, biological systems are adaptive by nature. With the shift toward human-relevant *in vitro* models and *in silico* modeling, it is more pressing than ever to differentiate between an adaptive response and an adverse perturbation of a biological system.

For both drug and chemical safety, the cost of gathering data is astounding – $20 billion in annual spending for safety assessments, with more on regulation, yet often this money is spent generating data with methods that were introduced 50 years ago (Meigs et al., 2018), or worse still, arguing about the significance of a tumor in one or two rodents (Williams et al., 2016; Davoren and Schiestl, 2018). Previous failures in this paradigm – thalidomide is a classic example – have been dealt with by requiring more animal testing, rather than thinking about what knowledge-gap led to this outcome. The precise mechanisms behind thalidomide's developmental toxicity have taken over 60 years to unravel. A crucial clue came from the observation that a mutation in the SALL4 gene, a suspected molecular target, results in a rare human condition resembling thalidomide toxicity. Structural differences between the human and mouse SALL4 proteins explain the species-specific differences in thalidomide toxicity (Donovan et al., 2018).

At the same time, data has increased in scale and diversity. Bringing ProbRA to the molecular level will no doubt require integrating multiple data streams at diverse scales – *in silico* ADME and docking studies, biological networks and pathways modeling, *in chemico* assays, high throughput and high dimensional data such as Toxcast and omics data, all the way to population-level exposure and outcome measurements. However, knowledge gained from one scale can inform the others, as has been demonstrated by the use of *in silico* metabolite predictions to help de-noise exposomics data (Kincaid et al., 2023).

It will be a challenge to efficiently and transparently integrate all these disparate data, and it will require an investment in data ontologies and AI-assisted transparent data integration techniques to ensure regulators have confidence in the output. Larger data sets do not always guarantee that the signal will be stronger than the noise. Paraphrasing Nate Silver, we must accept that distinguishing the signal from the noise requires scientific knowledge but also self-knowledge. We should have the courage to predict the endpoints that we can and the humility to accept that there are endpoints we cannot (yet) predict – and ideally, the wisdom to know the difference (Silver, 2012).

8.1 Data availability and needs for advancing probabilistic risk assessment

The advancement of ProbRA in toxicology relies heavily on the availability and quality of data. While the amount of toxicological data has been increasing rapidly, there are still significant challenges in terms of data accessibility, standardization, and relevance to human health. This chapter explores the current state of data availability, the types of data needed, and the steps required to bridge the gap between available and desired data for ProbRA.

Challenges and opportunities of data availability are central to this transition. Large datasets are crucial for standardization and the development of robust probabilistic models. However, several challenges need to be addressed to ensure the representativeness and quality of the data:

1. *Population bias:* If we assume that the available data represents a representative sample of the population, we may be skewed. Exposures are often higher in groups that are less likely to share data, leading to bias in the population structure.
2. *Open access:* A significant portion of the scientific literature is hidden behind paywalls. Major open access publishers, such as PLOS and Frontiers, are among the top sources of information for AI models like ChatGPT. To effectively use data science to address societal problems, including toxicological risk assessment, it is essential to promote open access to research data while ensuring the publication market is affordable for publishing authors.
3. *Data types:* The availability of different data types, such as pathology, omics, and electronic health records, is crucial. In defining the mission of ProbRA, it is important to consider what data types are available and can be integrated into the models in the short term.

One of the key challenges in advancing ProbRA is bridging the gap between *in vitro* systems, such as MPS or organoids, and human data. Efforts are needed to define how to translate findings from these advanced *in vitro* models to human health outcomes. Standardization initiatives, such as the MIAME guidelines for reporting the quality of transcriptomics data (Brazma et al., 2001), can help ensure the comparability and reliability of data from different sources.

Historically, regulatory agencies have relied heavily on animal data, but there is a growing interest in moving towards human-relevant data for risk assessment, as reflected in the OECD guidelines. Developing a vision for ProbRA based on human data requires exploring innovative approaches to leverage existing and emerging data sources, such as exposomics and multi-omics analysis of human blood and urine samples.

8.2 Generating relevant data for probabilistic risk assessment

To enable the development of ProbRA models, it is important to generate relevant and high-quality data. Lessons learned from initiatives like ToxCast 3.0 can help guide future efforts and avoid pitfalls, such as inadequate dose-response modeling or failing to account for batch effects.

Exposomics (Sillé et al., 2020), combined with biomonitoring, represents a promising approach to measure exposures from all sources, ultimately enabling a Human Exposome Project (Hartung, 2023a; Sillé et al., 2024). High-throughput omics technologies can provide valuable information on the observed perturbations in biological systems. Toxicogenomics data from liver cells or zebrafish, for example, can be compared with omics data from human blood samples to identify relevant pathways and mechanisms.

In summary, for the way forward to advance ProbRA, it is essential to:

1. Bring data together from diverse sources and make it operable
2. Promote open access and remove data from behind paywalls
3. Establish exposomics biobanks to support comprehensive exposure analysis
4. Generate high-content omics data from relevant model systems

Additionally, it is important to shift the focus from initiating events to understanding the long-term consequences of chemical exposures. By considering the complex dynamics and multiple descriptors of biological systems, including genetic factors, we can develop more accurate and predictive ProbRA models.

In conclusion, advancing ProbRA requires a concerted effort to address data availability, accessibility, and relevance. By promoting open access, leveraging innovative data sources, and generating high-quality data from human-relevant model systems, we can bridge the gap between available and desired data. Collaboration between regulatory agencies, industry, and academia is essential to overcome the challenges and realize the full potential of ProbRA in protecting human health.

9 Moving chemical risk assessment towards probabilistic approaches

Chemical risk assessment is a vital tool for ensuring the safety of chemicals in food and other applications, thereby protecting public health. As the field of toxicology evolves, there is a growing interest in moving from traditional deterministic approaches to probabilistic methods that better capture the inherent uncertainties and variabilities in chemical risk assessment. This chapter explores the principles of chemical risk assessment, the current use of probabilistic approaches, and the key questions that need to be addressed to further advance the field.

Probabilistic approaches are increasingly being used in chemical risk assessment to better characterize uncertainties and variabilities. Some examples of current applications include:

1. *Benchmark dose (BMD) modeling:* Used to relate biomarkers of exposure and effect to adverse outcomes, as demonstrated in the 2009 assessment of cadmium in food (EFSA, 2009).
2. *Physiologically-based pharmacokinetic (PBPK) modeling:* Combined with *in vitro*-to-*in vivo* extrapolation (IVIVE) to predict the distribution of a chemical in the human population, accounting for isoform-specific differences in metabolism.
3. *Modeling applications:* Used to address kinetic variability, perform population-specific simulations, and derive PODs based on internal dose, target organ or subcellular targets.

To further advance the use of probabilistic approaches in chemical risk assessment, several key questions need to be addressed:

1. *Fit-for-purpose approach:* Is a one-size-fits-all approach appropriate, or should the approach be tailored to the specific question being addressed?
2. *Expressing outcomes:* How can outcomes such as mutagenicity be expressed as probabilities rather than binary (yes/no) results?
3. *Incorporating NAMs:* How can NAMs, such as *in silico* methods, be integrated into ProbRA, and what considerations are needed regarding uncertainty factors, quantitative *in vitro*-to-*in vivo* extrapolation (QIVIVE), and the sensitivity of different endpoints?

Collaboration and data sharing are essential for advancing probabilistic approaches in chemical risk assessment. Initiatives such as joint databases and research projects between regulatory agencies, like ECHA and the OECD, can help to structure toxicity data and promote the use of NAMs for classification and prioritization.

In conclusion, probabilistic approaches offer a promising way forward for chemical risk assessment, enabling a more comprehensive characterization of uncertainties and variabilities. By addressing key questions related to fit-for-purpose approaches, expressing outcomes as probabilities, and incorporating NAMs, the field can continue to evolve and improve the protection of public health. Collaboration between regulatory agencies, industry, and the scientific community is crucial to ensure the successful implementation of probabilistic approaches in chemical risk assessment.

References

Alexandrov, L. B., Kim, J., Haradhvala, N. J. et al. (2020). The repertoire of mutational signatures in human cancer. *Nature 578*, 94-101. doi:10.1038/s41586-020-1943-3

Alhossary, A., Handoko, S. D., Mu, Y. et al. (2015). Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics 31*, 2214-2216. doi:10.1093/bioinformatics/btv082

Ames, B. N. and Gold, L. S. (1990). Misconceptions on pollution and the causes of cancer. *Angew Chem Int Ed Engl 29*, 1197-1208. doi:10.1002/anie.199011971

Ankley, G. T., Bennett, R. S., Erickson, R. J. et al. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem 29*, 730-741. doi:10.1002/etc.34

Aranda-Anzaldo, A. and Dent, M. A. R. (2018). Landscaping the epigenetic landscape of cancer. *Prog Biophys Mol Biol 140*, 155-174. doi:10.1016/j.pbiomolbio.2018.06.005

Bachman, J. A., Gyori, B. M. and Sorger, P. K. (2023). Automated

assembly of molecular mechanisms at scale from text mining and curated databases. *Mol Syst Biol 19*, e11325. doi:10.15252/msb.202211325

Basketter, D. A., Clewell, H., Kimber, I. et al. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX 29*, 3-89. doi: 10.14573/altex.2012.1.003

Bjørk, M.-H., Zoega, H., Leinonen, M. K. et al. (2022). Association of prenatal exposure to antiseizure medication with risk of autism and intellectual disability. *JAMA Neurol 79*, 672-681. doi:10.1001/jamaneurol.2022.1269

Blum, A., Behl, M., Birnbaum, L. et al. (2019). Organophosphate ester flame retardants: Are they a regrettable substitution for polybrominated diphenyl ethers? *Environ Sci Technol Lett 6*, 638-649. doi:10.1021/acs.estlett.9b00582

Bønnelykke, K., Sparks, R., Waage, J. et al. (2015). Genetics of allergy and allergic sensitization: Common variants, rare mutations. *Curr Opin Immunol 36*, 115-126. doi:10.1016/j.coi.2015.08.002

Bordukova, M., Makarov, N., Rodriguez-Esteban, R. et al. (2024). Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. *Expert Opin Drug Discov 19*, 33-42. doi:10.1080/17460441.2023.2273839

Bouhifd, M., Hogberg, H. T., Kleensang, A. et al. (2014). Mapping the human toxome by systems toxicology. *Basic Clin Pharmacol Toxicol 115*, 1-8. doi:10.1111/bcpt.12198

Bouhifd, M., Andersen, M. E., Baghdikian, C. et al. (2015). The human toxome project. *ALTEX 32*, 112-124. doi:10.14573/altex.1502091

Brazma, A., Hingamp, P., Quackenbush, J. et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet 29*, 365-371. doi:10.1038/ng1201-365

Calabrese, E. J. (2005). Paradigm lost, paradigm found: The re-emergence of hormesis as a fundamental dose response model in the toxicological sciences. *Environ Pollut 138*, 379-411. doi:10.1016/j.envpol.2004.10.001

Celardo, I., Aschner, M., Ashton, R. S., et al. (2025). Developmental neurotoxicity (DNT): A call for implementation of new approach methodologies for regulatory purposes: Summary of the 5th International Conference on DNT Testing. *ALTEX 42*, 323-349. doi:10.14573/altex.2503191

Chandrasekaran, S. N., Cimini, B. A., Goodale, A. et al. (2024). Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nat Methods 21*, 1114-1121. doi:10.1038/s41592-024-02241-6

Chernoff, J. (2021). The two-hit theory hits 50. *Mol Biol Cell 32*, rt1. doi:10.1091/mbc.e21-08-0407

Christensen, J., Grønborg, T. K., Sørensen, M. J. et al. (2013). Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *Surv Anesthesiol 57*, 292-293. doi:10.1097/01.sa.0000435536.89926.0a

Cöllen, E., Bartmann, K., Blum, J. et al. (2025). Mapping out strategies to further develop human-relevant, new approach methodology (NAM)-based developmental neurotoxicity (DNT) testing. *ALTEX 42*, 204-223. doi:10.14573/altex.2410112

Committee on the Design and Evaluation of Safer Chemical Substitutions (2014). *Overview of the GHS Classification Scheme in Hazard Classification*. Washington, DC, USA: National Academies Press.

Corradi, M., Luechtefeld, T., de Haan, A. M. et al. (2024). The application of natural language processing for the extraction of mechanistic information in toxicology. *Front Toxicol 6*, 1393662. doi:10.3389/ftox.2024.1393662

Coupland, C. A. C., Hill, T., Dening, T. et al. (2019). Anticholinergic drug exposure and the risk of dementia: A nested case-control study. *JAMA Intern Med 179*, 1084-1093. doi:10.1001/jamainternmed.2019.0677

Cozzini, P., Cavaliere, F., Spaggiari, G. et al. (2022). Computational methods on food contact chemicals: Big data and in silico screening on nuclear receptors family. *Chemosphere 292*, 133422. doi:10.1016/j.chemosphere.2021.133422

Craig, M., Gevertz, J. L., Kareva, I. et al. (2023) A practical guide for the generation of model-based virtual clinical trials. *Front Syst Biol 3*, 1174647. doi:10.3389/fsysb.2023.1174647

Dankovic, D. A., Naumann, B. D., Maier, A. et al. (2015). The scientific basis of uncertainty factors used in setting occupational exposure limits. *J Occup Environ Hyg 12, Suppl 1*, S55-68. doi:10.1080/15459624.2015.1060325

Davoren, M. J. and Schiestl, R. H. (2018). Glyphosate-based herbicides and cancer risk: A post-IARC decision review of potential mechanisms, policy and avenues of research. *Carcinogenesis 39*, 1207-1215. doi:10.1093/carcin/bgy105

de Bruyn Kops, C., Šícho, M., Mazzolari, A. et al. (2021). GLORYx: Prediction of the metabolites resulting from phase 1 and phase 2 biotransformations of xenobiotics. *Chem Res Toxicol 32*, 286-299. doi:10.1021/acs.chemrestox.0c00224

de Vries, R. B. M., Angrish, M., Browne, P. et al. (2021). Applying evidence-based methods to the development and use of adverse outcome pathways construct mechanistic frameworks for the development and use of non-animal toxicity tests. *ALTEX 38*, 336-347. doi:10.14573/altex.2101211

Debad, S. J., Aungst, J., Carstens, K. et al. (2025). State of the science on assessing developmental neurotoxicity using new approach methods. *ALTEX 42*, 121-144. doi:10.14573/altex.2410231

Dehaan, R. L. (1971). Toxicity of tissue culture media exposed to polyvinyl chloride plastic. *Nat New Biol 231*, 85-86. doi:10.1038/newbio231085a0

Djoumbou-Feunang, Y., Fiamoncini, J., Gil-de-la-Fuente, A. et al. (2019). BioTransformer: A comprehensive computational tool for small molecule metabolism prediction and metabolite identification. *J Cheminform 11*, 2. doi:10.1186/s13321-018-0324-5

Donovan, K. A., An, J., Nowak, R. P. et al. (2018). Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane Radial Ray syndrome. *Elife 7*, e38430. doi:10.7554/eLife.38430

EFSA (2009). Scientific opinion of the panel on contaminants in the food chain on a request from the European Commission on cadmium in food. *EFSA J 7*, 980. doi:10.2903/j.efsa.2009.980

EFSA (2021). Development of integrated approaches to testing and assessment (IATA) case studies on developmental neuro-

toxicity (DNT) risk assessment. *EFSA J 19*, 6599. doi:10.2903/j.efsa.2021.6599

Flynn, D. (2019). No cancer warnings required on coffee sold in California after all. *Food Safety News*.

Gao, Y., Mughal, Z., Jaramillo-Villegas, J. A. et al. (2024). BioBricks.Ai: A versioned data registry for life sciences data assets. *ArXiv.* doi:10.48550/arXiv.2408.17320

Gaylor, D. W. (2005). Are tumor incidence rates from chronic bioassays telling us what we need to know about carcinogens? *Regul Toxicol Pharmacol 41*, 128-133. doi:10.1016/j.yrtph.2004.11.001

Golden, E., Ukaegbu, D. C., Ranslow, P. et al. (2023). The good, the bad, and the perplexing: Structural alerts and read-across for predicting skin sensitization using human data. *Chem Res Toxicol 36*, 734-746. doi:10.1021/acs.chemrestox.2c00383

Golden, E., Allen, D., Amberg, A. et al. (2024). Toward implementing virtual control groups in nonclinical safety studies. *ALTEX 41*, 282-301. doi:10.14573/altex.2310041

Gottmann, E., Kramer, S., Pfahringer, B. et al. (2001). Data quality in predictive toxicology: Reproducibility of rodent carcinogenicity experiments. *Environ Health Perspect 109*, 509-514. doi:10.1289/ehp.01109509

Grabowska, M. E., Chun, B., Moya, R. et al. (2021). Computational model of cardiomyocyte apoptosis identifies mechanisms of tyrosine kinase inhibitor-induced cardiotoxicity. *J Mol Cell Cardiol 155*, 66-77. doi:10.1016/j.yjmcc.2021.02.014

Grimstein, M., Yang, Y., Zhang, X. et al. (2019). Physiologically based pharmacokinetic modeling in regulatory science: An update from the U.S. food and Drug Administration's Office of Clinical Pharmacology. *J Pharm Sci 108*, 21-25. doi:10.1016/j.xphs.2018.10.033

Gyori, B. M., Bachman, J. A., Subramanian, K. et al. (2017). From word models to executable models of signaling networks using automated assembly. *Mol Syst Biol 13*, 954. doi:10.15252/msb.20177651

Hanahan, D. (2022). Hallmarks of cancer: New dimensions. *Cancer Discov 12*, 31-46. doi:10.1158/2159-8290.cd-21-1059

Hartung, T. and McBride, M. (2011). Food for thought… on mapping the human toxome. *ALTEX 28*, 83-93. doi:10.14573/altex.2011.2.083

Hartung, T. (2016). Making big sense from big data in toxicology by read-across. *ALTEX 33*, 83-93. doi:10.14573/altex.1603091

Hartung, T., FitzGerald, R., Jennings, P. et al. (2017). Systems toxicology – Real world applications and opportunities. *Chem Res Toxicol 30*, 870-882. doi:10.1021/acs.chemrestox.7b00003

Hartung, T. (2023a). A call for a human exposome project. *ALTEX 40*, 4-33. doi:10.14573/altex.2301061

Hartung, T. (2023b). ToxAIcology – The evolving role of artificial intelligence in advancing toxicology and modernizing regulatory science. *ALTEX 40*, 559-570. doi:10.14573/altex.2309191

Hartung, T. (2023c). Artifical intelligence as the new frontier in chemical risk assessment. *Front Artif Intell 6*, 1269932. doi:10.3389/frai.2023.1269932

Hartung, T. and Kleinstreuer, N. (2025). Challenges and opportunities for validation of AI-based new approach methods. *ALTEX 42*, 3-21. doi:10.14573/altex.2412291

Hartung, T. and Smirnova, L. (2025). A path forward advancing microphysiological systems. *ALTEX 42*, 183-203. doi:10.14573/altex.2504091

Hemedan, A. A., Schneider, R. and Ostaszewski, M. (2023). Applications of Boolean modeling to study the dynamics of a complex disease and therapeutics responses. *Front Bioinform 3*, 1189723. doi:10.3389/fbinf.2023.1189723

Hemedan, A. A., Satagopam, V., Schneider, R. et al. (2024). Cohort-specific boolean models highlight different regulatory modules during Parkinson's disease progression. *iScience 27*, 110956. doi:10.1016/j.isci.2024.110956

Hoffmann, S., de Vries, R. B. M., Stephens, M. L. et al. (2017). A primer on systematic reviews in toxicology. *Arch Toxicol 91*, 2551-2575. doi:10.1007/s00204-017-1980-3

Jentink, J., Loane, M. A., Dolk, H. et al. (2010). Valproic acid monotherapy in pregnancy and major congenital malformations. *N Engl J Med 362*, 2185-2193. doi:10.1056/NEJMoa0907328

Jumper, J., Evans, R., Pritzel, A. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 596*, 583-589. doi:10.1038/s41586-021-03819-2

Kafka, P. (1998). Observations on risk management policies, focusing on experiences from their implementation and use in the field of nuclear technology. In: *Proceedings of ESA Risk Management Workshop, Noordwijk* (85-100). European Space Agency, ESA-ESTEC.

Kincaid, B., Piechota, P., Golden, E. et al. (2023). Using in silico tools to predict flame retardant metabolites for more informative exposomics-based approaches. *Front Toxicol 5*, 1216802. doi:10.3389/ftox.2023.1216802

King, C., Fowler, J. C., Abnizova, I. et al. (2023). Somatic mutations in facial skin from countries of contrasting skin cancer risk. *Nat Genet 55*, 1440-1447. doi:10.1038/s41588-023-01468-x

Kleensang, A., Maertens, A., Rosenberg, M. et al. (2014). t^4 workshop report: Pathways of toxicity. *ALTEX 31*, 53-61. doi:10.14573/altex.1309261

Kleensang, A., Vantangoli, M. M., Odwin-DaCosta, S. et al. (2016). Genetic variability in a frozen batch of MCF-7 cells invisible in routine authentication affecting cell function. *Sci Rep 6*, 28994. doi:10.1038/srep28994

Kleinstreuer, N. and Hartung, T. (2024). Artificial Intelligence (AI) – It's the end of the tox as we know it (and I feel fine) – AI for predictive toxicology. *Arch Toxicol 98*, 735-754. doi:10.1007/s00204-023-03666-2

Knudson, A. G. (1996). Hereditary cancer: Two hits revisited. *J Cancer Res Clin Oncol 122*, 135-140. doi:10.1007/BF01366952

Krewski, D., Acosta, D., Jr, Andersen, M. et al. (2010). Toxicity testing in the 21st century: A vision and a strategy. *J Toxicol Environ Health B Crit Rev 13*, 51-138. doi:10.1080/10937404.2010.483176

Leist, M., Ghallab, A., Graepel, R. et al. (2017). Adverse outcome pathways: Opportunities, limitations and open questions. *Arch Toxicol 31*, 221-229. doi:10.1007/s00204-017-2045-3

Letort, G., Montagud, A., Stoll, G. et al. (2019). PhysiBoSS: A multi-scale agent-based modelling framework integrating physical dimension and cell signalling. *Bioinformatics 35*, 1188-1196.

doi:10.1093/bioinformatics/bty766

Liao, Z., You, R., Huang, X. et al. (2019). DeepDock: Enhancing ligand-protein interaction prediction by a combination of ligand and structure information. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, USA, 311-317. doi:10.1109/bibm47256.2019.8983365

Linkov, I., Massey, O., Keisler, J. et al. (2015). From "weight of evidence" to quantitative data integration using multicriteria decision analysis and Bayesian methods. *ALTEX 32*, 3-8. doi:10.14573/altex.1412231

Loizou, G. D. (2016). Animal-free chemical safety assessment. *Front Pharmacol 7*, 218. doi:10.3389/fphar.2016.00218

Luechtefeld, T., Maertens, A., McKim, J. M. et al. (2015). Probabilistic hazard assessment for skin sensitization potency by dose-response modeling using feature elimination instead of quantitative structure-activity relationships. *J Appl Toxicol 35*, 1361-1371. doi:10.1002/jat.3172

Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016a). Analysis of Draize eye irritation testing and its prediction by mining publicly available 2008-2014 REACH data. *ALTEX 33*, 123-134. doi:10.14573/altex.1510053

Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016b). Global analysis of publicly available safety data for 9,801 substances registered under REACH from 2008-2014. *ALTEX 33*, 95-109. doi:10.14573/altex.1510052

Luechtefeld, T. and Hartung, T. (2017). Computational approaches to chemical hazard assessment. *ALTEX 34*, 459-478. doi:10.14573/altex.1710141

Luechtefeld, T., Marsh, D., Rowlands, C. et al. (2018a). Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility. *Toxicol Sci 165*, 198-212. doi:10.1093/toxsci/kfy152

Luechtefeld, T., Rowlands, C. and Hartung, T. (2018b). Big-data and machine learning to revamp computational toxicology and its use in risk assessment. *Toxicol Res 7*, 732-744, doi:10.1039/c8tx00051d

Maertens, A., Luechtefeld, T., Kleensang, A. et al. (2015). MPTP's pathway of toxicity indicates central role of transcription factor SP1. *Arch Toxicol 89*, 743-755. doi:10.1007/s00204-015-1509-6

Maertens, A., Bouhifd, M., Zhao, L. et al. (2017). Metabolomic network analysis of estrogen-stimulated MCF-7 cells: A comparison of overrepresentation analysis, quantitative enrichment analysis and pathway analysis versus metabolite network analysis. *Arch Toxicol 91*, 217-230. doi:10.1007/s00204-016-1695-x

Maertens, A., Golden, E., Luechtefeld, T. H. et al. (2022). Probabilistic risk assessment – The keystone for the future of toxicology. *ALTEX 39*, 3-9. doi:10.14573/altex.2201081

Marx, U., Andersson, T.B., Bahinski, A. et al. (2016). Biology-inspired microphysiological system approaches to solve the prediction dilemma of substance testing using animals. *ALTEX 33*, 272-321. doi:10.14573/altex.1603161

Marx, U., Akabane, T., Andersson, T. B. et al. (2020). Biology-inspired microphysiological systems to advance medicines for patient benefit and animal welfare. *ALTEX 37*, 364-394. doi:10.14573/altex.2001241

Marx, U., Beken, S., Chen, Z. et al. (2025). Biology-inspired dynamic microphysiological system approaches to revolutionize basic research, healthcare and animal welfare. *ALTEX 42*, 204-223. doi:10.14573/altex.2410112

McNally, K. and Loizou, G. D. (2015). A probabilistic model of human variability in physiology for future application to dose reconstruction and QIVIVE. *Front Pharmacol 6*, 213. doi:10.3389/fphar.2015.00213

Meigs, L., Smirnova, L., Rovida, C. et al. (2018). Animal testing and its alternatives – The most important omics is economics. *ALTEX 35*, 275-305. doi:10.14573/altex.1807041

Montagud, A., Béal, J., Tobalina, L. et al. (2022). Patient-specific Boolean models of signalling networks guide personalised treatments. *Elife 11*, e72626 doi:10.7554/eLife.72626.sa2

Musuamba, F. T., Skottheim Rusten, I., Lesage, R. et al. (2021). Scientific and regulatory evaluation of mechanistic in silico drug and disease models in drug development: Building model credibility. *CPT Pharmacometrics Syst Pharmacol 10*, 804-825. doi:10.1002/psp4.12669

Niarakis, A., Ostaszewski, M., Mazein, A. et al. (2023). Drug-target identification in COVID-19 disease mechanisms using computational systems biology approaches. *Front Immunol 14*, 1282859. doi:10.3389/fimmu.2023.1282859

O'Hara, T., Virág, L., Varró, A. et al. (2011). Simulation of the undiseased human cardiac ventricular action potential: Model formulation and experimental validation. *PLoS Comput Biol 7*, e1002061. doi:10.1371/journal.pcbi.1002061

Osimitz, T. G., Droege, W., Boobis, A. R. et al. (2013). Evaluation of the utility of the lifetime mouse bioassay in the identification of cancer hazards for humans. *Food Chem Toxicol 60*, 550-562. doi:10.1016/j.fct.2013.08.020

Paller, C. J., Tukachinsky, H., Maertens, A. et al. (2024). Pan-cancer interrogation of MUTYH variants reveals biallelic inactivation and defective base excision repair across a spectrum of solid tumors. *JCO Precis Oncol 8*, e2300251. doi:10.1200/PO.23.00251

Pamies, D. and Hartung, T. (2017). 21st century cell culture for 21st century toxicology. *Chem Res Toxicol 30*, 43-52. doi:10.1021/acs.chemrestox.6b00269

Parreno, V., Loubiere, V., Schuettengruber, B. et al. (2024). Transient loss of Polycomb components induces an epigenetic cancer fate. *Nature 629*, 688-696. doi:10.1038/s41586-024-07328-w

Passini, E., Britton, O. J., Lu, H. R. et al. (2017). Human in silico drug trials demonstrate higher accuracy than animal models in predicting clinical pro-arrhythmic cardiotoxicity. *Front Physiol 8*, 668. doi:10.3389/fphys.2017.00668

Place, T. L., Domann, F. E. and Case, A. J. (2017). Limitations of oxygen delivery to cells in culture: An underappreciated problem in basic and translational research. *Free Radic Biol Med 113*, 311-322. doi:10.1016/j.freeradbiomed.2017.10.003

Reynolds, E. H. and Green, R. (2020). Valproate and folate: Congenital and developmental risks. *Epilepsy Behav 108*, 107068. doi:10.1016/j.yebeh.2020.107068

Ross, B. C., Boguslav, M., Weeks, H. et al. (2018). Simulating heterogeneous populations using Boolean models. *BMC Syst Biol 12*, 64. doi:10.1186/s12918-018-0591-9

Sillé, F. C. M., Karakitsios, S., Kleensang, A. et al. (2020). The exposome – A new approach for risk assessment. *ALTEX 37*, 3-23. doi:10.14573/altex.2001051

Sillé, F. C. M., Busquet, F., Fitzpatrick, S. et al. (2024). The implementation moonshot project for alternative chemical testing (IMPACT) toward a human exposome project. *ALTEX 41*, 344-362. doi:10.14573/altex.2407081

Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail – But Some Don't*. Penguin.

Sluka, J. P., Fu, X., Swat, M. et al. (2016). A liver-centric multiscale modeling framework for xenobiotics. *PLoS One 11*, e0162428. doi:10.1371/journal.pone.0162428

Smirnova, L., Hogberg, H. T., Leist, M. and Hartung, T. (2014). Deve-lopmental neurotoxicity – challenges in the 21st century and in vitro opportunities. *ALTEX 31*, 129-156. doi:10.14573/altex.1403271

Smirnova, L., Kleinstreuer, N., Corvi, R. et al. (2018). 3S – Systematic, systemic, and systems biology and toxicology. *ALTEX 35*, 139-162. doi:10.14573/altex.1804051

Smirnova, L., Hogberg, H. T., Leist, M. and Hartung, T. (2024). Revolutionizing developmental neurotoxicity testing – a journey from animal models to advanced in vitro systems. *ALTEX 41*, 152-178. doi:10.14573/altex.2403281

Suciu, I., Pamies, D., Peruzzo, R. et al. (2023). G × E interactions as a basis for toxicological uncertainty. *Arch Toxicol 97*, 2035-2049. doi:10.1007/s00204-023-03500-9

Sun, J. X., He, Y., Sanford, E. et al. (2018). A computational approach to distinguish somatic vs. germline origin of genomic alterations from deep sequencing of cancer specimens without a matched normal. *PLoS Comput Biol 14*, e1005965. doi:10.1371/journal.pcbi.1005965

Tomasetti, C. and Vogelstein, B. (2015). Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science 347*, 78-81. doi:10.1126/science.1260825

Tomczak, K., Czerwińka, P. and Wiznerowicz, M. (2015). The cancer genome atlas (TCGA): An immeasurable source of knowledge. *Contemp Oncol 19*, A68-77. doi:10.5114/wo.2014.47136

Tran, V., Kim, R., Maertens, M. et al. (2021). Similarities and differences in gene expression networks between the breast cancer cell line Michigan cancer foundation-7 and invasive human breast cancer tissues. *Front Artif Intell 4*, 674370. doi:10.3389/frai.2021.674370

Trosko, J. E. and Upham, B. L. (2005). The emperor wears no clothes in the field of carcinogen risk assessment: Ignored concepts in cancer risk assessment. *Mutagenesis 20*, 81-92. doi:10.1093/mutage/gei017

Trott, O. and Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem 31*, 455-461. doi:10.1002/jcc.21334

Vertrees, R. A., Jordan, J. M., Solley, T. et al. (2009). Tissue culture models. In T. C. Allen and P. T. Cagle (eds), *Basic Concepts of Molecular Pathology* (159-82). Boston, MA, USA: Springer US. doi:10.1007/978-0-387-89626-7_18

Vinken, M., Benfenati, E., Busquet, F. et al. (2021). Safer chemicals using less animals: Kick-off of the European ONTOX project. *Toxicology 458*, 152846. doi:10.1016/j.tox.2021.152846

Weindling, P. (1985). *The Social History of Occupational Health*. Croom Helm.

Williams, G. M., Aardema, M., Acquavella, J. et al. (2016). A review of the carcinogenic potential of glyphosate by four independent expert panels and comparison to the IARC assessment. *Crit Rev Toxicol 46*, 3-20. doi:10.1080/10408444.2016.1214677

Conflict of interest

None declared.

Data availability

No datasets were generated in this study.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 963845 (ONTOX). We would like to thank Dr Bas Bokkers, National Institute for Public Health and the Environment (RIVM), The Netherlands, for valuable discussion.