OXFORD
SOT | Society of Toxicology
academic.oup.com/toxsci

Toxicological Sciences, 2023, **192(2),** 155–177
https://doi.org/10.1093/toxsci/kfad012
Advance Access Publication Date: February 14, 2023
In-Depth Review

A pragmatic framework for the application of new approach methodologies in one health toxicological risk assessment

Kelly A Magurany,[1,*] Xiaoqing Chang,[2] Rebecca Clewell,[3] Sandra Coecke,[4] Esther Haugabrooks,[5] Sue Marty[6]

[1]NSF International, Ann Arbor, Michigan 48105, USA
[2]Inotiv-RTP, Morrisville, North Carolina 27560, USA
[3]21st Century Tox Consulting, Chapel Hill, North Carolina 27517, USA
[4]European Commission Joint Research Centre, Ispra, Italy
[5]Coca-Cola Company (formerly Physicians Committee for Responsible Medicine), Atlanta, Georgia 30313, USA
[6]The Dow Chemical Company, Midland, Michigan 48667, USA

*To whom correspondence should be addressed at NSF International, 789 N. Dixboro Rd, Ann Arbor, MI 48105, USA. E-mail: kmagurany@nsf.org

Abstract

Globally, industries and regulatory authorities are faced with an urgent need to assess the potential adverse effects of chemicals more efficiently by embracing new approach methodologies (NAMs). NAMs include cell and tissue methods (*in vitro*), structure-based/toxicokinetic models (*in silico*), methods that assess toxicant interactions with biological macromolecules (*in chemico*), and alternative models. Increasing knowledge on chemical toxicokinetics (what the body does with chemicals) and toxicodynamics (what the chemicals do with the body) obtained from *in silico* and *in vitro* systems continues to provide opportunities for modernizing chemical risk assessments. However, directly leveraging *in vitro* and *in silico* data for derivation of human health-based reference values has not received regulatory acceptance due to uncertainties in extrapolating NAM results to human populations, including metabolism, complex biological pathways, multiple exposures, interindividual susceptibility and vulnerable populations. The objective of this article is to provide a standardized pragmatic framework that applies integrated approaches with a focus on quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) to extrapolate *in vitro* cellular exposures to human equivalent doses from which human reference values can be derived. The proposed framework intends to systematically account for the complexities in extrapolation and data interpretation to support sound human health safety decisions in diverse industrial sectors (food systems, cosmetics, industrial chemicals, pharmaceuticals etc.). Case studies of chemical entities, using new and existing data, are presented to demonstrate the utility of the proposed framework while highlighting potential sources of human population bias and uncertainty, and the importance of Good Method and Reporting Practices.

Keywords: new approach methodologies (NAMs); risk assessment; framework; metabolism; uncertainty; *in vitro*

The discovery, development and marketing of new chemistries (eg, cosmetics, pharmaceuticals, industrial or consumer products, and food supplements) continue to impact growing segments of the global industrial production market (Deconinck, 2021; Pandey *et al.*, 2022). This may raise concerns as the World Health Organization (WHO) reported in 2021 that more than 1 million global deaths were attributable to chemical exposures (World Health Organization (WHO), 2021). Further, recognition of the interconnectedness of the health of people, animals, plants and the environment within the One Health concept speaks to the need for a multifaceted and collaborative chemical strategy (Bronzwaer *et al.*, 2022; Center for Disease Control (CDC), 2022; World Health Organization (WHO), 2022). To improve public safety and address current global One Health perceptions (de Jongh *et al.*, 2022), the European Commission published its chemicals strategy for sustainability towards a toxic-free environment (the "strategy") as part of the European Green Deal (European Commission (EC), 2019, 2020). It presents several actions to bring about a toxic-free environment and to protect people and the environment from hazardous chemicals. Similarly, in March of 2020, the US Environmental Protection Agency published their Chemical Safety for Sustainability Strategic Research Action plan for 2019–2022 in an effort "*…to support safe selection, design and use of chemicals and materials*" (United States Environmental Protection Agency (US EPA), 2020a). Globally, several additional health and food production strategies across industrial sectors aim to act in line with the toxic-free movement to identify, replace, restrict or ban carcinogenic, mutagenic and reprotoxic substances (CMRs), endocrine disruptors, persistent, bioaccumulative and toxic (PBT) and very persistent and very bioaccumulative (vPvB) substances, immunotoxicants (including respiratory sensitizers), neurotoxicants, and substances toxic to specific organs (DG SANTE, 2022; Joint FAO/WHO Expert Committee on Food Additives (JECFA), 2022; Joint FAO/WHO Meeting on

Pesticide Residues (JMPR), 2022; United States Environmental Protection Agency (US EPA), 2022a; World Health Organization (WHO), 2021).

To accommodate risk assessment and management needs for a variety of industrial and consumer product sectors, conventional risk assessment practices using animal studies cannot keep pace with demands for modern risk evaluation as *in vivo* rodent studies require extensive time to conduct studies and report results. They are also expensive and have low throughput (United States Environmental Protection Agency, Office of Research and Development (US EPA), 2015). Although animal studies have been effective in evaluating some aspects of chemical safety, additional limitations include the ethics associated with testing chemicals in animals, the large number of animals required to test each substance (>1000), the limited information provided on mechanisms of observed effects and the inability to capture human genetic variability in response (acknowledging human homologous cell lines may also be limited in capturing genetic variability at this time) (United States Environmental Protection Agency, Office of Research and Development (US EPA), 2015; United States Food and Drug Administration (US FDA), 2021a). Furthermore, although many physiological processes are conserved between humans and animals, animal experiments may not translate to humans due to differences in design, execution and analysis of animal experiments or species differences when compared to either human clinical trials or chemical real-life exposure scenarios. These deficiencies are important because animal research often provides the rationale for hypotheses studied by epidemiologists and clinical researchers. Furthermore, translational relevance to humans for a variety of adverse health effects has been questioned, including reproductive and endocrine toxicity (Schenk *et al.*, 2010; Solecki *et al.*, 2019), developmental neurotoxicity (Bal-Price, 2018), carcinogenic modes of action (Cohen *et al.*, 2019), and the transition of new pharmaceuticals to Phase III clinical trials (that have also been limited in interhuman predictivity based on postmarket toxicity) (Parish *et al.*, 2020). In addition, interspecies kinetic differences (ie, absorption, distribution, metabolism, and excretion [ADME]) also limit animal to human extrapolation (Pelkonen, 2009; Punt, 2017). When toxicology studies in animal species do not fully capture the human *in vivo* mechanistic-related adverse events (eg, receptor expression patterns or binding affinities) or the human or individual diversity in adverse health effects, results from *in vivo* animal studies may under or over predict toxicity in humans (Elmeliegy *et al.*, 2021). Because new chemical entities continue to enter the market, there is some urgency for effective management of chemical safety with applications of NAMs to fill gaps for chemicals lacking sufficient safety data or comprehensive risk evaluations.

NAMs in toxicology, that include any technology that provides safety and risk assessment data (eg, *in silico*, *in vitro*, *in chemico, or* alternative animal models) (Parish *et al.*, 2020; Punt 2020; United States Environmental Protection Agency, Office of Research and Development (US EPA), 2015), provide an opportunity to advance the science of risk assessment with human (individual) relevant, mechanism-based, and higher throughput methods at lower costs. Alignment of NAM development with the WHO/IPCS Mode of Action Framework (Meek *et al.*, 2014), the OECD Adverse Outcome Pathway (AOP) Knowledge-Base (Organization for Economic Co-operation and Development (OECD), 2014), and the OECD Defined Approach NAM-based guidelines (Organization for Economic Co-operation and Development (OECD), 2021a) facilitates the development of methods consistent with our understanding of key events (KEs) in the biological pathways that mediate known adverse human health effects (Meek and Lipscomb, 2015; Parish *et al.*, 2020). (See Figure 1 for an illustration of AOP and related terms.) Such approaches are critical to support the need for shorter timelines and greater efficacy for safe chemicals management by assisting regulatory authorities to supplement animal study data, move away from traditional animal testing that is not aligned with the current state of science, or employ innovative approaches to model complex chemical interactions using systems biology (European Union (EU), 2010; Novak *et al.*, 2022; United States Environmental Protection Agency (US EPA), 2019; United States Food and Drug Administration (US FDA)'s, 2021b). This approach was envisioned in 'One Health', which will use NAMs, physiologically based kinetic (PBK) modeling, and measured outcomes (if and when they occur) to assess risk factors and predict outcomes in the presence of complex stressors (eg, pollution, environmental degradation, climate change, etc.) (Center for Disease Control (CDC), 2022; World Health Organization (WHO), 2022).

Broad adoption of NAMs has been limited partially due to insufficient validation, standardization, and global harmonization of *in vitro*, *in silico*, alternative animal and *in chemico* new approach methods for regulatory use. *In vitro* new approach methodologies can be simple cellular, subcellular or *in chemico* methods or complex 3D methods in microphysiological systems (eg, tissue-, organ-, or body-on-a-chip methods). As such, the experimental design, experimental biokinetics, measurement methods and interpretation require qualified personnel with specialized technical knowledge. This poses challenges in coordinating global efforts to align and accelerate method development and acceptance (Krebs *et al.*, 2020; Pain *et al.*, 2020). In addition, uncertainties and a standard process for evaluation of bias and compliance with minimal good cell/tissue culture practices need to be addressed, including transparent criteria and a robust process to evaluate method readiness and the derived study data (Organization for Economic Co-operation and Development (OECD), 2018a; Coecke *et al.*, 2016). Although NAMs are not fully developed for all systemic endpoints, NAM data derived from the published literature clearly contribute to our understanding of human kinetics and dynamics (Bal-Price, 2018; Krebs *et al.*, 2020; Parish *et al.*, 2020). Furthermore, regulatory authorities are responding to the challenge of using NAMs to assess hazards and risks to human and ecosystem health (Barton-Maclaren *et al.*, 2022). In addition, NAM data are being used more frequently for prioritization, to add to the weight-of-evidence (WoE) for regulatory decision-making, or to fill data gaps from animal testing, including estimating equivalent human administered doses (EADs) from *in vitro* assay data using quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) (Bell *et al.*, 2018; Casey *et al.*, 2018; Wilk-Zasadna *et al.*, 2015).

To support further NAM utilization in risk assessment, this publication provides a decision framework (see Figure 2) for the risk assessment community with a standardized procedure to apply NAMs while accounting for system complexities and areas of uncertainty. The goal of the framework is to simplify the application of NAMs by providing a defined process and an explanation of the associated tools and resources for each decision point. Although *in vitro* methods are the focus of the framework, it is expected that *in silico and in chemico* techniques may be applied within several decision points, eg, estimation of physicochemical and metabolic parameters that would inform QIVIVE and/or prediction of the mode of action (MoA) or AOP based on similar chemistries (Pradeep *et al.*, 2020a,b; Tennant *et al.*, 2019; Zang

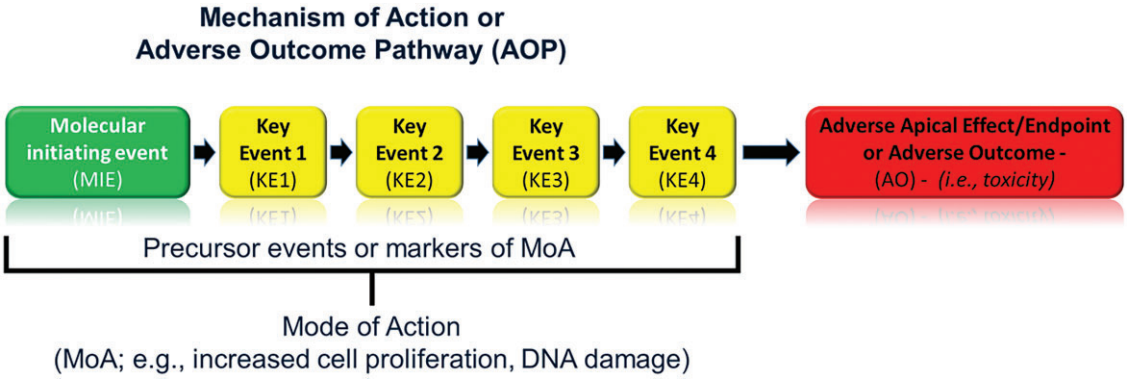

Figure 1. Depicted here is the basic relationship of the mode of action (MoA) that describes the functional or anatomical change that leads to an adverse apical effect or adverse outcome (AO), and the mechanism of action, or adverse outcome pathway (AOP), that describes the specific biomolecular and/or biochemical events that occur within the MoA. The adverse apical effect, or AO, describes the observed toxicity in the exposed organism that is the result of a chemical acting through the mode and mechanism of action. The molecular initiating event (MIE) is the initial biomolecular interaction that leads to an AO and each key event that follow describes the subsequent steps within the pathway leading to the AO. (Adapted from Felter et al. (2022). See also Clerbaux et al. (2022) for an illustration of the AOP.)

et al., 2017). In addition, case studies are provided to demonstrate the application of the framework. To further inform NAM usage, additional case studies are available in the published literature (eg, Firman *et al.*, 2021; Gannon *et al.*, 2019; Hartwig *et al.*, 2020; Ning *et al.*, 2019; Organization for Economic Co-operation and Development (OECD), 2020, 2022a; Ouedraogo *et al.*, 2022; Paul Friedman *et al.*, 2020; Ramanarayanan *et al.*, 2022; Rotroff *et al.*, 2013; Rovida, 2021). Lastly, although the framework is expected to be broadly applicable, special cases including nanoparticles, polymers, poorly soluble chemicals, chemical mixtures, and medical devices require additional considerations that are not covered as example case studies in this article, and thus, readers are directed to specific literature for additional guidance on the application of NAMs for these materials. (For examples of NAMs application for special case studies, see Tirumala *et al.* (2021) and Halappanavar *et al.* (2020) for nanomaterials, Schüttler *et al.* (2021) for chemical mixtures, Ladics *et al.* (2021) for poorly soluble, polymeric materials, and De Jong *et al.* (2020) for medical devices).

The framework, tools, and available resources

The framework is designed as an iterative decision tree providing key decision points in the application of *in vitro*-based NAMs in risk assessment. Ultimately, the outcome is to derive margins of safety (MoS) for chemical exposures or reference values based on points of departure identified from *in vitro* methods, using either forward or reverse kinetic models, respectively. The framework is not intended to imply that all NAMs are suitable for regulatory decision-making but is intended to facilitate the use of fit-for-purpose NAM techniques to identify comparative points of departure or reference values to inform human risk evaluations in the context of a WoE approach. In this way, NAM data may be interpreted in the appropriate context considering current uncertainties associated with their predictive value relative to the *in vivo* condition. As NAM technology advances, the predictive quality of these approaches will also advance, and this framework may provide the tools needed for ready application and adoption by the general risk assessment community.

Each colored pane of the framework identifies the major tasks of any risk assessment and includes (1) framing the assessment, (2) evaluating chemical dynamics and bioactivity, (3) *in vitro* method interpretation and, finally, (4) kinetics in the context of identifying estimated reference doses (RfDs) or MoSs for known chemical exposure from *in vitro* points of departure. The framework has been intentionally designed to be generic for broad applicability to multiple risk assessment contexts. In support of this framework and to advance the understanding and adoption for the use of NAMs in risk assessment, the available tools and resources for each decision point in the process are provided. Note, this list of resources is an attempt to collect the relevant information that has been identified for each of these decision points but may not represent a comprehensive list. Although the framework intends to provide a simplified decision tree on a KEs basis, additional endpoints, and AOPs and associated KEs may be relevant and thus utilization of the framework may be repeated within any one assessment. As a single KE is also unlikely to describe the complex biology *in vivo* for any apical outcome, a battery of assays representing multiple KEs may need to be considered to provide a sufficient WoE to determine risk. At present, there is not a comprehensive battery of assays that covers all organ systems and thus use of a solely NAMs-based approach is not possible. However, this does not limit the value of existing NAMs data for investigated endpoints and their incorporation into a WoE-based risk assessment. Further, other methods in risk assessment that promote effective utilization of resources, such as read across of data on similar chemicals, may also be applied in lieu of generating new NAM data as suggested in this framework.

It is important to recognize that the presented framework is not the first framework published to support the use and confidence in NAMs. However, this framework is the first that presents a holistic overview of the necessary decision points and associated resources for the application of NAMs for the purposes of identifying reference values or margins of safety (MoS) in the risk assessment context. Other frameworks, for example, by Parish *et al.* (2020) and van der Zalm *et al.* (2022), provide guidance to determine whether a method is appropriate and scientifically robust for the context of use, whether it be for prioritization, screening or risk assessment and whether the method appropriately informs an understanding of human biology and the mechanism of toxicity in the *in vivo* system. These frameworks would therefore inform decision points 2.6 and 3.2 of the presented framework and are thus incorporated as resources for the same. In addition, both frameworks describe the importance of problem

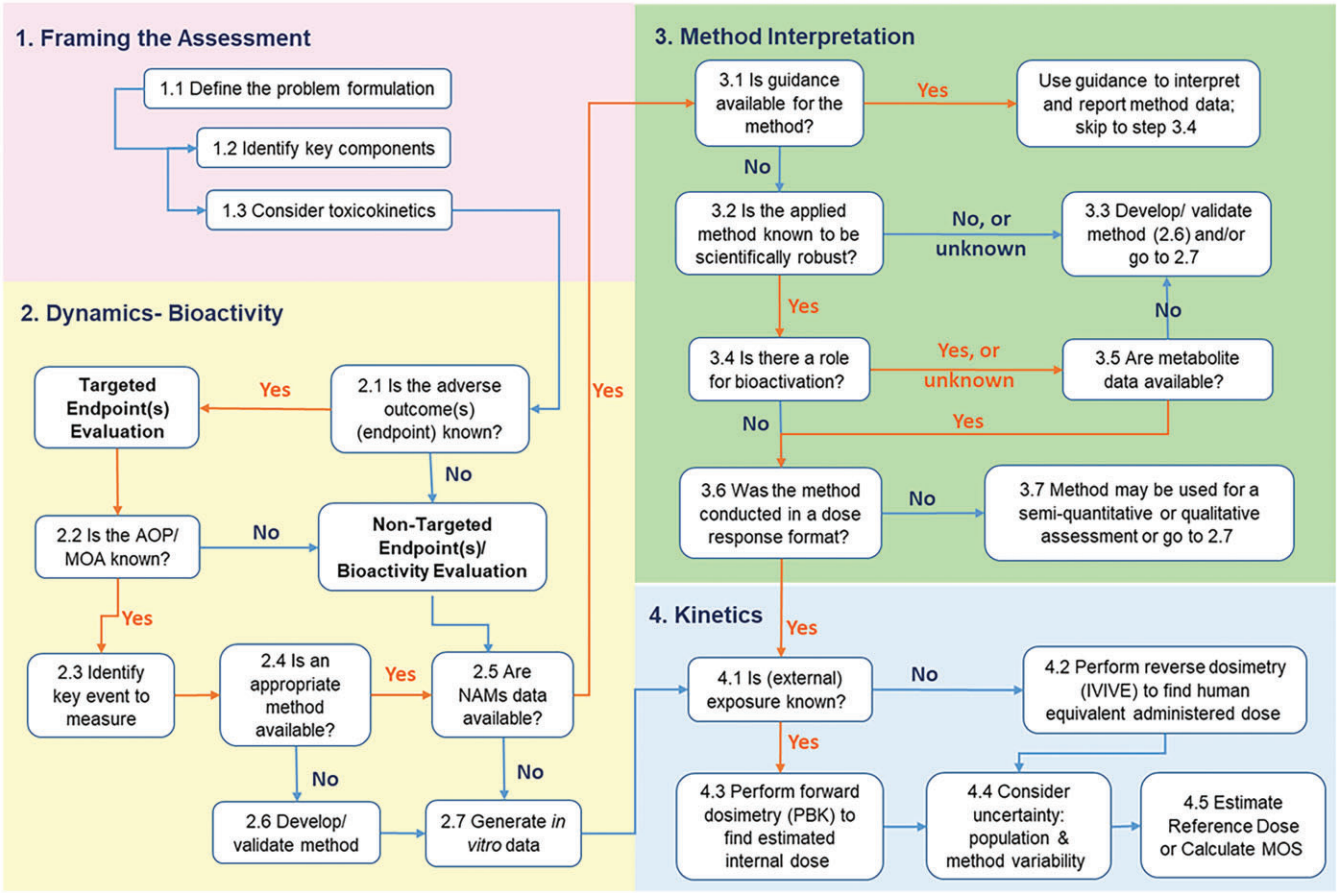

Figure 2. A proposed framework for the application of *in vitro* new approach methodologies (NAMs) in risk assessment is provided. Four key components for considerations in use of NAMs are described including (1) Framing the Assessment which outlines the problem formulation, key components of the assessment and considers the toxicokinetics of the subject compound as might influence the use or interpretation of NAMs; (2) Dynamics-Bioactivity which considers the apical effect, or adverse outcome, of the chemical and whether a targeted or nontargeted approach may be applied. In cases where the adverse outcome and its associated pathway is understood and *in vitro* methods are available to measure the key event(s), a targeted approach may be applied. In cases where the adverse outcome and its associated pathway is not understood, a nontargeted approach may be applied that broadly investigates *in vitro* bioactivity (eg, using toxicogenomics). The resulting bioactivity information may be used to investigate the adverse outcome pathway and identify a key event to measure or these data may be used directly to identify a conservative point of departure based solely on bioactivity. (3) Method Interpretation provides steps to evaluate method results considering the availability and validity of scientific methods and whether the available data are sufficient to characterize internal dosimetry relative to chemical bioactivation *in vivo* (ie, is the toxic moiety of the chemical known and is the *in vitro* method metabolically competent such that the target tissue dose of the active compound is represented in the method?); Finally, in step (4) Kinetics, where *in vitro* method data have been determined to be scientifically robust, relevant to target tissue concentrations of the active toxic moiety and run in a dose-response format, the human equivalent administered dose (EAD) or internal dose may be estimated from the *in vitro* data or external exposure using physiologically based kinetic (PBK) models. Human equivalent external or internal doses estimated from PBK models may then be used to identify an estimated reference dose (RfD) to establish a level of exposure that would be without appreciable risk to human or environmental health for the relevant endpoint or based absence of bioactivity, or to allow for calculation of a margin of safety (MoS) for those cases where exposure is known. Both the RfD and MoS approaches should consider the uncertainty associated with population and method variability and the estimated external and internal doses should be adjusted using associated uncertainty factors, as applicable. Further guidance for each decision point is provided in the text of the article and in the Supplementary Table S-1.

formulation and thus also inform decision point 1.1 of this framework. As such and as intended, the presented framework serves as a basic structure in which further guidance for the associated decisions points may be developed to foster and further inform NAMs confidence and implementation into risk assessment.

Framing the assessment

Problem formulation (1.1)

Problem formulation is the most critical component of the risk assessment and defines the purpose of the assessment, wherein the problem is defined based on what is known and what can be adapted from what is known and outlines a plan and conceptual model for analyzing and characterizing hazard or risk, as applicable. Problem formulation has been well described in the available literature and these resources for developing the problem formulation can be found in Supplementary Table S-1. Good examples of formal problem formulation documents include those developed within the US EPA for chemicals undergoing risk evaluation under the Toxic Substances Control Act (TSCA) (United States Environmental Protection Agency (US EPA), 2020b).

Identify key components (1.2)

This step identifies key information, parameters or tools that are needed to execute the problem formulation plan. This requires an assessment of available data, identification of data needs and "fit for purpose" NAMs, and consideration of variables that can minimize uncertainty in the assessment.

Consider toxicokinetics (1.3)

In vitro methods that characterize a hazardous concentration in a target tissue for a defined adverse outcome or a specific KE in an AOP, rarely consider the influence of the absorption, distribution, metabolism, or elimination (ADME) of a chemical in the body on *in vivo* toxicity. As such, the target tissue concentration *in vivo* may not represent *in vitro* conditions (ie, may over or under predict toxicity) or the putative toxicant may be absent in the *in vitro* system if a metabolite is active. However, to ensure that a NAM-based approach is fit for purpose for an assessment as defined in the problem formulation, toxicokinetics should be considered early in the assessment process. Interpretation of NAM data for *in vivo* risk requires testing with chemical concentrations similar to those expected in plasma or target tissue and ensuring relevance of the test methods to the route of exposure, endpoint of interest, and the active chemical moiety (parent vs metabolite(s)) (Moreau *et al.*, 2022).

Dosimetry is important for both hazard identification (ID) and risk assessment to provide confidence that the *in vitro* method has been tested under the appropriate conditions. Even in hazard ID, which is considered a qualitative assessment, there is typically an *in vivo* limit dose above which it is not considered necessary to continue testing. Currently, most *in vitro* high throughput screening efforts set an upper limit concentration based on solubility or cytotoxicity cutoff. ToxCast phase 1 used an upper limit of 100 μM. However, depending on the chemical, this may or may not be sufficient to observe effects. Many well-known carcinogens, for example, do not cause effects until they are in the mM range (eg, methylmethane sulfonate; Doak *et al.*, 2007). Thus, a negative result at 100 μM would present as a false negative. To ensure that *in vitro* testing conditions are suitable for the purpose of the assessment, toxicokinetics should be used up-front to help define the assay conditions.

Likewise, while bioactivation/inactivation is inherently accounted for in animal studies (species differences notwithstanding), *in vitro* systems do not typically account for metabolism when measuring bioactivity. However, considering toxicokinetics during study design can help ensure that the method accounts for the contribution of metabolism to overall toxicity, whether the goal is risk assessment or hazard ID. Where data are not available, read across to metabolism for similar chemicals or utilization of metabolism simulators within Organization for Economic Co-operation and Development (OECD) Toolbox v 4.5 (2021b) may be used to predict potential chemical metabolites.

The availability and regulatory acceptance of NAMs to assess the cell and tissue dynamics and kinetics (ADME/biokinetics) in chemical risk evaluations has been a critical area of research for over a decade (Blaauboer, 2010; Coecke *et al.*, 2006; Rotroff *et al.*, 2010). A range of possibilities for the use of NAMs for biokinetics in risk evaluations were formulated (eg, to define species differences, human variability or to perform quantitative *in vitro* to *in vivo* extrapolations; Testai et al., 2021).To increase the regulatory use and acceptance of NAMs for dynamics and biokinetics, the development of test guidelines (protocols) and of overarching guidance documents is critical and are actively being developed (see Supplementary Table S-2 for published examples). A key document published by Coecke et al. (2006) as an outcome of ECVAM's 54th workshop is still relevant today that describes the metabolism "bottleneck" that is present in the use of *in vitro* methods. Research in recent years also has identified challenges defining *in vitro* exposure levels for some test materials (ie, free fraction) and the need to avoid crossover dosing to adjacent wells with semivolatile compounds (eg, Birch et al., 2019).

Dynamics-bioactivity

Is the adverse outcome(s) (endpoint) known? (2.1)

Is the AOP or one or a limited number of modes of action (MoA) known? (2.2)

Identify KE to measure (2.3)

The adverse outcome or endpoint is the biologically significant effect that a chemical exerts on a living organism or population that results in impairment of functional capacity (Keller et al., 2012; Organization for Economic Co-operation and Development (OECD), 2017a). Interpretation of NAM data characterizing adverse effects versus adaptive responses (that do not result in functional deficits) is critical to adequately characterize health risk. The problem formulation determines the scope of the adverse endpoint(s) that will be considered in the assessment and may be focused on a targeted endpoint (eg, estrogen receptor agonism) or a broad range of possible adverse outcomes in a nontargeted assessment.

AOP and mechanism or MoA are interrelated terms (see Figure 1) that describe the etiology of the observed or potential toxicity. KEs that are known to occur prior to the observation of the apical effect provide points of evaluation using NAMs to predict the likelihood for toxicity in the absence of animal data. Pathways of toxicity are actively being mapped and may be described within the OECD AOP knowledge base (https://aopkb.oecd.org/) or may be published in the peer-reviewed literature. Examples in the published literature include Barron et al. (2015) for aquatic toxicity, Browne et al. (2017) for endocrine disruption, and Cohen *et al.* (2019) and Hartwig et al. (2020) for carcinogenicity and genotoxicity. Additional guidance documents for the investigation of AOPs or chemical MoAs are provided in Supplementary Table S-1.

Is an appropriate method available? (2.4)

The most appropriate method will depend on the problem formulation and the associated endpoint of interest (ie, a targeted adverse outcome and KEs or a nontargeted bioactivity concentration-response assessment). Ideally, when considering data for risk assessment, methods would be validated, having available OECD test guidelines; however, the scope of NAMs having OECD guidelines or defined approaches are limited and thus, establishing fit-for-purpose may be sufficient and the appropriateness of a NAM should be evaluated in the context of the assessment. Some sources of methodological data and resources for evaluating fit-for-purpose are listed in Supplementary Table S-1.

Are NAM data available? (2.5)

NAM data are available for many chemicals based on the extensive testing efforts conducted around the world (some global programs are listed in Supplementary Table S-2). Supplementary Table S-1 may be used to identify data in addition to specific chemical reports that may be found in the published literature. In addition, NAM data for similar chemicals within a read-across approach may be considered. In these cases, good read across practices (GRAP) should be applied (Ball et al., 2016).

Develop/validate method (2.6)

Best practices for *in vitro* methodologies are well-outlined in the resources highlighted in Supplementary Table S-1. It is of critical importance that the method is in line with good cell and tissue culture practices (Pamies et al., 2021) for developing *in vitro* methods and approaches for scientific and regulatory use. E-learning modules also are available that provide guidance to test method developers (eg, see https://etplas.eu/learn/eu-60/).

For methodologies that are not fully validated, additional resources to evaluate study quality also are available in Supplementary Tables S-1, S-3, and S-4 and are important as it is recognized that many *in vitro* test methods that are not fully validated may be fit-for-purpose for regulatory application (Bal-Price, 2018).

Generate NAM data (2.7)

Ideally, *in vitro* data should be generated using validated, where possible, or robust scientific methods with the evaluation conducted in a dose-response format to allow for identification of the point of departure to establish safe human exposure levels. However, as previously mentioned, methods that are not validated may be fit-for-purpose and should be evaluated as such within the context of use. See Supplementary Table S-1 for decision point 2.4 for resources on how to identify an appropriate method and 2.6 for key resources associated with method validation and guidelines for the use of nonvalidated methods.

The internationally recognized Organization for Economic Co-operation and Development (OECD) (2018) guidance document on Good *in vitro* Method Practice (GIVIMP) is intended to support method developers and end-users working to establish new *in vitro* methods in academic, industry or government laboratories across all 38 OECD member countries, and beyond, to increase the quality of the studies carried out with such methods, and to reduce experimental bias as much as possible. In Figure 3, a series of *in vitro* subcell, cell, and tissue OECD GIVIMP minimal reporting essentials are depicted. In the area of compound safety, GIVIMP serves as a comprehensive quality framework for the development or execution of *in vitro* methods. GIVIMP and other select approaches to ensure NAM quality and control bias are listed in Supplementary Table S-1.

Auditing tools are also provided in the supplementary material to assess systematically the methodological information necessary to interpret more completely *in vitro* study data. OECD issued a specific OHT201 template (Organization for Economic Co-operation and Development (OECD), 2021c) that can be completed in compliance with GCCP (good cell culture practice) as part of GIVIMP when *in vitro* mechanistic NAMs are reported. The template (Supplementary Table S-3) offers the possibility to declare that GIVIMP guidance was followed and illustrates a synopsis of *in vitro* subcell, cell, and tissue models with minimal reporting essentials when generating *in vitro* NAM data using the variety of experimental set ups. Supplementary Table S-4 indicates potential sources of bias that should be considered.

Test data derived from *in vitro* methods are increasingly being used in combination with other information within Integrated Approaches to Testing and Assessment (IATA) to support safety decisions. It is strongly recommended to take careful consideration of GIVIMP requirements during the development of *in vitro* methods as this will help improve the quality of submitted methods, accelerate their acceptability for regulatory use and ultimately reduce the experimental bias on the derived *in vitro* method study data. In short, any *in vitro* study should provide information about the cells or tissues used (test system) to arrive at the measurements, the detection method, the method used for dose selection, control, and reference chemicals used, specific experimental conditions, data analysis, acceptance criteria applied, validity of the data, reporting of results and uncertainties.

Method interpretation

Is guidance available for the method? (3.1) *See answer to 2.4 above*

Is the applied method known to be scientifically robust? (3.2) *See answer to 2.6 above*

Develop/validate method or generate data (3.3). *See 2.6 above*

Is there a role for bioactivation? (3.4)

The vast majority of *in vitro* methods focus on parent chemical effects, with little consideration of metabolism in toxicity. When metabolism is considered, it is typically assumed to be inactivating and is incorporated into *in vitro* assessments through inclusion of parent chemical clearance in the data interpretation phase of the study. In reality, many toxic chemicals require bioactivation to exert toxic effects including many that cause thyroid hormone disruption (Murk et al., 2013) and a majority of carcinogens (Miller and Miller, 1975; Pelkonen and Vähäkangas, 1980). Rendic and Peter Guengerich (2012) suggested that about 66% of environmental carcinogens require cytochrome P450 enzyme-mediated bioactivation. Currently, the most common approach to incorporating bioactivation into an *in vitro* method is adding rat liver S9 fraction to the cell culture system. This is a fairly well-established approach to genetic toxicity testing and recently has been the focus of an effort by the US EPA Center for Computational Toxicology and Exposure (CCTE) to incorporate S9 fractions into high throughput screening (Deisenroth et al., 2020). Limitations to this approach include the use of rat liver, the use of induced S9 which leads to high expression of a subset of enzymes, and the potential toxicity of the S9 itself on the cells of interest. Other approaches that are being explored include the use of QSAR models to predict metabolism and bioactivity of metabolites (Gonzalez et al., 2018), and flow-through systems that connect metabolically competent hepatocytes to target cell populations (Wang et al., 2019). Bioreactors can be used to provide more realistic metabolite kinetic profiles than static systems (Phillips et al., 2018), though the added complexity of the system adds cost and time to the analyses. If the bioactive metabolite is known, it may be possible to obtain metabolites and directly test the metabolite in the system of interest (Clewell et al., 2020). There is no "one size fits all" approach for incorporating metabolism, but it is vitally important to consider whether metabolism—and bioactivation in particular—is likely to be required for a toxic response to avoid false negative results (Wilk-Zasadna et al., 2015). Predictive models of chemical metabolism such as those included in the OECD Toolbox v 4.5 or evaluation of metabolism and putative toxicants for similar chemical structures may be considered in the absence of metabolic data.

Are metabolite data available? (3.5) *See answer to 2.5 above*

Was the method conducted in a dose-response format? (3.6)

A point of departure used to derive a reference value or to establish a MoS cannot be effectively identified without the availability of dose-response data. From the dose-response curve, a statistically based

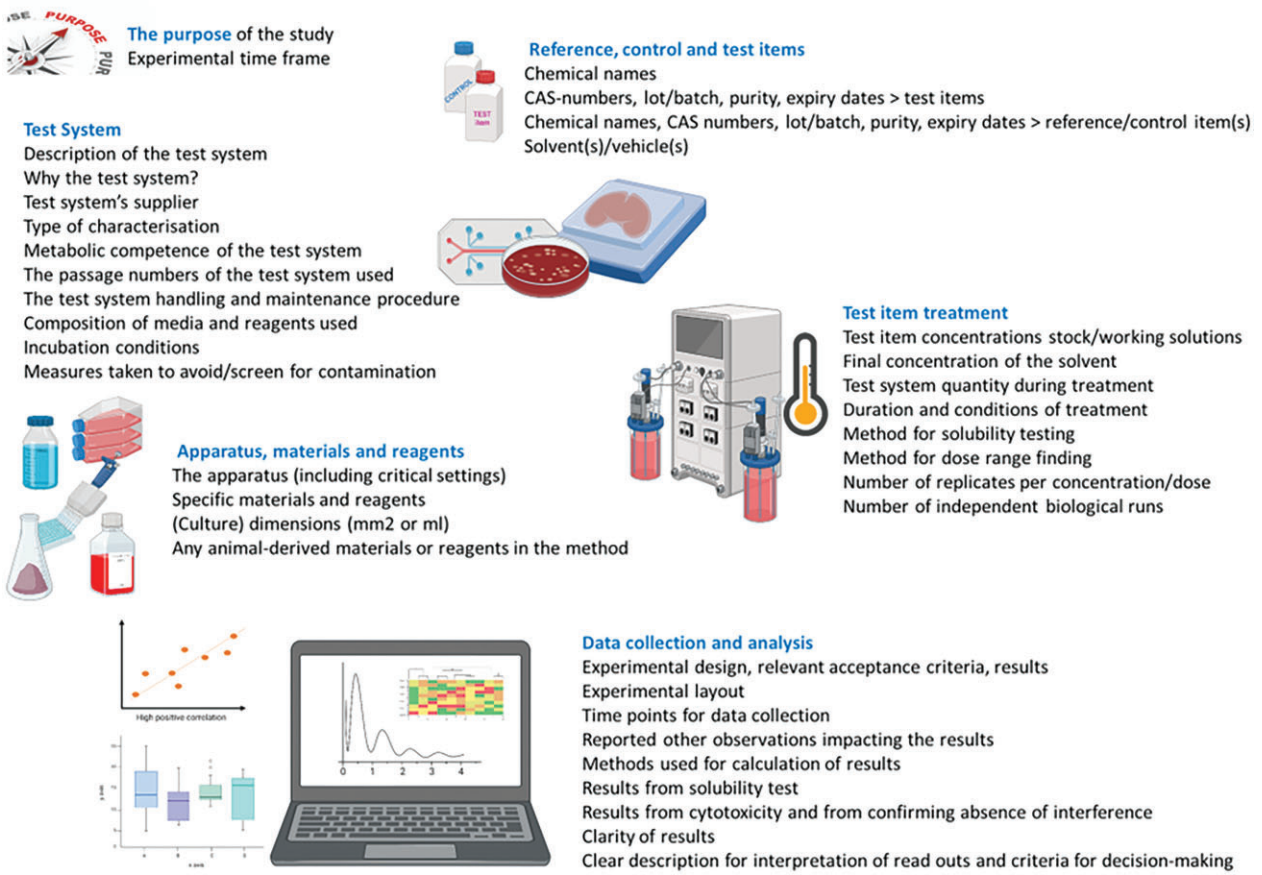

Fig 3. *In vitro* subcell, cell, and tissue OECD GIVIMP (2018b) minimal reporting essentials.

point of departure identified using benchmark dose modeling may be identified. In the context of an *in vitro* method as described in ToxCast, this may be an AC50 (activity concentration at 50% of the 100% maximal activity) or ACC (activity concentration at the cutoff of the method) value. The method cut-off is method specific; derived as baseline median absolute deviations (BMAD) from baseline (signal to noise), or log2 fold induction as compared to the vehicle control activity or the 2 lowest test concentrations of the method. Points of departure from other NAMs would need to be identified specific to the test system. For example, induction levels of genetic transcription would be relevant to toxicogenomic methods.

Method may be used for a semiquantitative or qualitative assessment or go to 2.7 (3.7)

In the absence of dose-response data, *in vitro* data may be used as part of the WoE in a semiquantitative or qualitative manner. Depending on the understanding of the AOP relative to apical effects, these data may potentially be used to support identification of a hazard or to characterize the likely MoA or activity for a KE in the AOP. At the very least, they may inform the direction for the development of dose-responsive NAM data.

Kinetics (PBK models and IVIVE)

Is (external) exposure known? (4.1)

In vivo kinetics describes the ADME of chemicals in the body and is important in determining the availability of a chemical at molecular target sites (Ruiz and Fowler, 2015). Sometimes exposures may be well characterized (eg, exposure monitoring) or critical exposure variables may be well defined (eg, cosmetics). In these cases, a physiologically based kinetic (PBK) model can be used to simulate ADME and predict internal (eg, plasma or tissue) concentration-time profiles of chemicals at the given external exposure. This process is referred to as forward dosimetry. In some cases, the external exposure is unknown, but measured biomonitoring data (eg, urine and blood concentration of parent compound or metabolites) are available. For these cases, a PBK model can be used to determine a plausible environmental exposure level corresponding to the biological monitoring data, which is referred to as reverse dosimetry. Reverse dosimetry is more and more commonly applied to environmental exposure evaluations (Wetmore et al., 2012).

An important consideration in selection and interpretation of a PBK model output is the dose metric that is being predicted by the model and the relevance to the chemical exposure and adverse effect. Dependent on the exposure scenario for the chemical (eg, acute vs chronic or cumulative exposure) and the adverse effect of interest, steady-state chemical concentration (Css), the maximum chemical concentration (C_{max}), or area under the concentration curve (AUC) may be selected as the appropriate dose metric.

QIVIVE: perform reverse dosimetry to find human equivalent administered dose (4.2)

As a broad definition, any utilization of *in vitro* measured bioactivity data to predict *in vivo* exposures can be referred to as *in vitro* to *in vivo* extrapolation, or IVIVE. IVIVE applies reverse dosimetry to relate an *in vitro* bioactivity concentration (instead of using biomonitoring data) to an *in vivo* exposure level that would lead to internal (plasma or tissue) concentration equal to

the *in vitro* bioactivity concentration. Thus, it is important to specify the context where IVIVE is applied. High-throughput (HT-)IVIVE, is the simplest approach to IVIVE, using only intrinsic clearance, glomerular filtration and fraction unbound in the blood to estimate steady state concentrations of a chemical in blood (Rotroff et al., 2010). Quantitative (Q)IVIVE (Yoon et al., 2015) extends this approach to consider more complex determinants of *in vivo* pharmacokinetics, including slow metabolism, bioactivation, active chemical transport, etc. This estimated external exposure level is often described as the equivalent administered dose (EAD) (Casey et al., 2018) or administered equivalent dose (AED), which can be compared to actual human exposures for risk evaluation or testing prioritization. IVIVE of dosimetry is important in putting *in vitro* method data into a relevant *in vivo* context and to promote animal-free risk assessment.

There are more and more resources and tools available for PBK modeling and conducting IVIVE (Breen et al., 2021; Chang et al., 2022). Madden *et al.* (2020) summarized *in silico* resources assisting construction and evaluation of PBK models, including resources for predicting external exposure, physicochemical properties, ADME properties, physiological parameters, models for specific organs structures, modeling software, etc. Chang *et al.* (2022) provided additional data sources and tools for PBK modeling as well as software and tools designed to support IVIVE of dosimetry. These resources include the Integrated Chemical Environment (ICE: https://ice.ntp.niehs.nih.gov/) that provides easy and open access to high-quality curated data and interactive PBK modeling and IVIVE tools to explore and contextualize *in vitro* bioactivity. ICE tools use the PBK models provided in the Httk R package (v 2.0.2, https://cran.r-project.org/web/packages/httk/index.html), which can also be run directly in R. The httk R package was developed by a group of scientists at US EPA (Pearce et al., 2017) and provides chemical-specific *in vitro* data, physiological information, and multiple functions for PBK modeling under various exposure routes and species as well as IVIVE of dosimetry. It also provides functions for evaluating uncertainty and variability in modeling (Ring et al., 2017).

Perform forward dosimetry (PBK modeling) to find the estimated internal dose (4.3)

PBK models simulate internal dose from a given external dose that considers use pattern, exposure frequency/amounts, uptake, etc. This is also referred to as forward dosimetry. In this case, internal dose (blood or tissue concentration) can be compared with an *in vitro* bioactive point of departure (PoD) to derive a MoS. A lower estimated internal dose (than the *in vitro* PoD) suggests the exposure scenario is very likely "safe" and shall not raise a concern, while a higher estimated internal dose (than the *in vitro* PoD) suggests the exposure is likely not "safe" and further testing is needed.

Consider uncertainty with NAMs and characterizing population variability in kinetics (4.4)

Uncertainty arises when data are lacking. Uncertainty with NAMs can be derived from assay interference, experimental artifacts, quantitative uncertainty in concentration-activity curve-fitting, uncertainty in QSAR modeling, *in vitro* and *in vivo* pharmacokinetic model parameterization (Pham et al., 2019). To consider uncertainty as the NAMs is being developed helps better design NAMs and builds confidence when applying NAMs in risk assessment. Uncertainty can be reduced or eliminated with more or better data. On the other hand, variability refers to the inherent heterogeneity or diversity of data, which can be quantitatively described as the range of a set of values (United States Environmental Protection Agency (US EPA), 2011). Compared to uncertainty, variability cannot be reduced, but it can be better characterized. Population variability describes how much a physiological or kinetic parameter varies within a given population. It is an important aspect when characterizing kinetics. The following provides a focused discussion on uncertainty and variability specifically associated with kinetics and IVIVE.

Use of uncertainty factors in traditional risk assessment

Uncertainty factors (UFs) have been used in traditional regulatory risk assessment to account for uncertainties when extrapolating toxicological responses from animal to human, average to sensitive human populations, lowest-observed-adverse-effect level (LOAEL) to no-observed-adverse-effect level (NOAEL), and from acute to chronic exposure scenarios. Whether uncertainty factors are needed when using human cell NAMs has been the subject of debate (Dourson et al., 2022). It has been suggested that the interspecies (animal to human) uncertainty factor of 10 could be eliminated when using data from *in vitro* human cells on a chip. But to eliminate or reduce the uncertainty factor in extrapolation from average human to sensitive human populations requires additional toxicokinetic and/or toxicodynamic information. Other standard uncertainty factors, such as subchronic to chronic extrapolation, might still be needed even when using *in vitro* human cells.

Sources of uncertainty in in vitro *dosimetry/bioactivity*

Confident application of NAMs requires the consideration of uncertainties and variability in bioactivity data generation, interpretation, and evaluation of the impact of these uncertainties on safety assessment conclusions. When interpreting IVIVE results, uncertainty analysis requires the identification of weaknesses in all facets of the scientific assessment and how these limitations could impact scientific conclusions (Organization for Economic Co-operation and Development (OECD), 2020; European Food Safety Authority (EFSA) *et al.*, 2019). Bioactivity uncertainty assessments should include several factors contributing to overall uncertainty for a specific NAM (Pham et al., 2019; Watt and Judson, 2018); for example: (1) NAM limitations to represent all important facets of critical cellular processes; (2) variability in *in vivo* data used to validate NAM performance; (3) remaining method gaps (eg, understanding assay precision/accuracy, variability, and reproducibility); (4) uncertainties in the positive and negative control chemicals used for *in vitro* assay validation; and (5) limitations common to many *in vitro* methods (eg, lack of metabolic competence, limited understanding of the domain of applicability, unknown *in vitro* kinetics). For example, when collecting data from *in vitro* systems, nominal concentrations are often used. However, nominal concentrations do not necessarily represent the cellular concentration or the free median concentration (Armitage et al., 2021).

Sources of variability and uncertainty in human toxicokinetic distributions and IVIVE

The sources of variability and uncertainty that need to be considered in IVIVE include: the *in vitro* active concentration, mechanistic relevance of NAMs to *in vivo* outcomes, interindividual variability in animal or human physiology, and uncertainty associated with measurement and estimation of chemical-dependent toxicokinetic parameters (eg, fraction unbound to protein, metabolic clearance, both measured and predicted; Wetmore et al.,

2015). Efforts have been given to address the uncertainly and variability associated with toxicokinetic parameters and population variability. Open-source tools, such as HTTK-Pop in Ring et al. (2017) examine the impact of variability of input parameters and population variability on estimated EAD values (Wambaugh et al., 2019).

Estimate reference dose or calculate the MoS (4.5)

In the case of reverse dosimetry, an RfD may be calculated by simply dividing the human equivalent administered dose generated from QIVIVE by the composite uncertainty factor, or EAD/Total UF = RfD. For forward dosimetry, a MoS is calculated by comparing the internal dose at which a chemical known to be bioactive or the *in vitro* activity concentration with the internal dose predicted from the known exposure using PBK models, expressed as the following equation:

$$\text{MoS} = \frac{\text{Known bioactive internal dose OR } \textit{in vitro} \text{ activity dose}}{\text{Estimated internal dose from given external exposure}}.$$

The acceptable value of the MoS may vary depending on the uncertainty in the data and the context of the risk assessment but is typically ≥ 100 (eg, Paul Friedman et al., 2020).

Case studies

NAMs and 'next generation risk assessments' require the integration of multiple data streams with greater consideration of exposure and underlying mechanisms of toxicity. The purpose of the framework is to put these concepts into practice, where mechanistic understanding is available, providing a path for data integration and highlighting points to consider at various steps within the process for greater consistency and transparency. The framework is flexible and therefore, widely applicable for a variety of chemical safety assessments. Three case studies are presented that illustrate the utility of the framework—benzophenone, N-methylmorpholine N-oxide (NMMO), and XU-18840.00. The rationale for selecting these case studies is provided in Table 1. Each case study focuses on integrating NAM data and QIVIVE to determine the risk posed by each chemical while considering the potential impact of bias and uncertainty during the evaluations.

Case study 1: Benzophenone (BP) in drinking water

A comprehensive risk assessment for BP was developed by NSF International and peer-reviewed by NSF International's independent Health Advisory Board (NSF International, 2021a). This assessment identified an appropriate RfD for BP that was used to derive acceptance criteria for drinking water where BP may extract from drinking water system components. An RfD of 0.02 mg/kg-day was identified based on the most sensitive adverse effect (ie, the critical effect) from a 2-year feeding study, which included increased syncytial alteration and chronic active inflammation of hepatocytes secondary to chronic enzyme induction in male B6C3F1 mice that occurred at a human equivalent LOAEL of 6.1 mg/kg-day. Splenic lymphoid follicular hyperplasia in male and female B6C3F1 mice and splenic cell proliferation in female B6C3F1 mice were also observed as critical effects at the LOAEL. Increased severity of chronic progressive nephropathy in rats occurred at a slightly higher but comparable point of departure.

Clear evidence of endocrine activity was not observed in a 2-generation reproductive toxicity study for BP (Hoshino et al., 2005; Yamasaki et al., 2005) until the high dose, where decreased testes weight and increased ovary weight were observed in F_0 male and female animals, which equates to a human equivalent dose ~4× higher than the observed critical effects for the liver and spleen. Notably, decreased anogenital distance (AGD) was observed in female F_1 offspring but not male offspring at the low-dose (−7%) and mid-dose (−10%) but not the high dose. AGD is a sensitive indicator of androgenic, anti-androgenic, and estrogenic effects, where a shortened AGD in male animals has been associated with genital malformations and decreased reproductive capacity into adulthood, and androgenic effects resulting in elongation of the AGD such as is observed in the masculinization of female offspring (Schwartz et al., 2019). Decreased AGD in female animals (single generation) in the absence of decreased AGD in male animals and other observation of endocrine effects is of unknown etiology and not currently characterized as clearly indicative of endocrine disruption, although it has been observed for other endocrine active compounds including ketoconazole, bisphenol A, and lindane (Schwartz et al., 2019). As such, a potential weakly estrogenic MoA to describe the decreased AGD in female animals cannot be completely discounted based on the weight of the evidence.

Risk assessment discussions relevant to the use of NAMs data and the framework to derive comparative RfDs for endocrine endpoints are provided in the following case study. For additional details on the full assessment including evaluation of all endpoints and available data, readers are directed to the NSF International (2021a) risk assessment.

Define the problem formulation (1.1); and

Identify key components (1.2)

The risk assessment of BP was predicated upon the need to derive acceptance criteria for exposure to residuals of BP, a UV stabilizer in coatings, that may extract into drinking water from materials certified for use in contact with drinking water according to NSF/ANSI/CAN 61 (2021b). The traditional aspects of the risk assessment were conducted in accordance with NSF/ANSI/CAN 600 (2021c) procedures and using traditional animal assays to identify the most sensitive point of departure. To evaluate the potential benefit of *in vitro* data for RfD derivation and to add to the WoE for investigation of endocrine endpoints, BP was selected as a case study to evaluate how the reference dose (RfD)derived from *in vitro* methods for endocrine endpoints and QIVIVE compared to the RfD derived from traditional animal data. The outcome of this assessment would inform whether the RfD derived from traditional animal assays would be protective of potential estrogen effects based on *in vitro* data characterizing the known MoA for estrogen agonism. Endocrine endpoints were identified as relevant for QIVIVE in this manner as the *in vitro* ToxCast battery had been validated for replacement of screening assays in animals for estrogen agonism and because there was an existing AOP that described effects from the receptor to organ (or apical effect) level (Browne et al., 2017). In addition, the criteria for inclusion in the assessment for this comparison included the availability of a 2-generation reproductive toxicity study and evidence of endocrine activity within the suite of *in vitro* endocrine methods evaluated within the US EPA ToxCast program. The ToxCast suite of methods for endocrine endpoints are described in detail by Judson et al. (2015) and are summarized in Supplementary Table S-5.

Table 1. Case study chemicals illustrating the utility of the NAM framework

Case Study Chemical	Chemical Uses	Toxicity	Selection Rationale
Benzophenone in drinking water	• UV stabilizer in coatings • Residual may extract from drinking water system components	• Reference dose (RfD) based on critical effects in mouse study (hepatocyte alterations and splenic lymphoid hyperplasia) • Potential endocrine activity not well characterized	• Targeted analysis with reverse dosimetry to determine the comparability of RfD derived from NAMs data and quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) for endocrine endpoints versus the RfDs derived from traditional animal data

N-methylmorpholine N-oxide (NMMO)	• Industrial solvent used to produce lyocell fiber (dissolves cellulose, which is reprecipitated to produce a fiber)	• Rats: Decreased sperm production and reduced fertility at >50 mg/kg/day
		• Goal:
		  • Evaluate human relevance of reproductive effects seen in rats
		  • If spermatotoxic, are humans more or less sensitive than rats
			• Targeted analysis with reverse dosimetry shows framework strengths to demonstrate species differences in:
			  • Pharmacodynamics based on rat vs monkey NAMs evaluating sperm maturational stages
			  • Pharmacokinetics wherein oral equivalent doses show spermatotoxicity is unlikely in monkeys and humans.

XU-18840.00	• New cosmetic ingredient for use in face creams	• Toxicity is not known as this is a new chemical; however, cosmetic ingredients cannot be tested on animals after 2013 if sold in the European Union per the Cosmetics Directive	• Nontargeted analysis to predict safety across a broad biological space. Safety determinations based on a MoS relative to the lowest point of departure (PoD) in the NAM assessment. • NAM PoD is compared via QIVIVE using forward dosimetry based on use patterns, concentration in face cream and dermal penetration through human skin.

Consider toxicokinetics (1.3)

The metabolic competency of the ToxCast suite of endocrine methods was partially described by Judson et al. (2015) and fully described in the United States Environmental Protection Agency (US EPA) (2022c) method documentation for non-guideline *in vitro* test methods. None of the methods, except for the Attagene methods (ATG_ERα_TRANS-UP; ATG_ERE_CIS_up), were metabolically competent; therefore, the endocrine activity of BP metabolites was considered to determine the feasibility of applying NAMs data for derivation of an RfD.

Based on a study by Nakagawa and Tayama (2002), a dose-dependent increase in the mean serum BP (~33%–45%), benzhydrol (BH) (51%–65%), *p*-hydroxybenzophenone (4-HBP) (~2%–3%) were identified 6 hours post-dosing in female Sprague Dawley rats with gavage doses of 100 and 400 mg/kg-day BP (>97% pure). These metabolites were also identified in the plasma of male Sprague Dawley rats after gavage dosing of 100 mg/kg-day BP (purity not specified) (Jeon et al., 2008). *In vivo* uterotrophic studies investigating the endocrine activity of BP and its metabolites suggest that *p*-hydroxybenzophenone may be a potent estrogen agonist (Nakagawa and Tayama, 2001; Yamasaki et al., 2002) and thus ToxCast data for both BP and its active metabolite, *p*-hydroxybenzophenone, were evaluated for comparative RfDs (see Table 2) (Benzhydrol was not considered endocrine active.)

Is the adverse outcome (endpoint) known? (2.1)

Yes. *In vitro* and *in vivo* data supporting evidence of endocrine activity, specifically estrogen receptor agonism, were available. As such, a targeted endpoint evaluation was conducted.

Is the AOP/MoA known? (2.2)

Yes. The AOP for estrogen receptor agonism is well described by Browne et al. (2017). Based on the AOP, it is expected that estrogen receptor binding and activation will result in apical effects that may include increased uterine and ovarian weight and altered histopathology at the organ level in uterotrophic methods (OECD TG 440) and in Level 5 endocrine studies (according to the Organization for Economic Co-operation and Development (OECD) (2012) Framework for the Testing and Assessment of Endocrine Disrupters) that include the extended one-generation reproductive toxicity study (OECD TG 443) and the 2-generation reproductive toxicity study (OECD TG 416).

Identify KE to measure (2.3)

The KE that was measured in the available *in vitro* methods was estrogen receptor agonism.

Is an appropriate method available? (2.4)

Yes. The ToxCast suite of *in vitro* estrogen-receptor assays have been validated as a replacement to the uterotrophic screening method (OECD TG 440) with a published AOP that associates the KE to apical effects (Browne et al., 2015, 2017).

Are NAMs data available? (2.5)

Yes. *In vitro* methods for the endocrine endpoints for BP and *p*-hydroxybenzophenone are available through multiple resources (National Center for Advancing Translational Science (NCATS), 2022; United States Environmental Protection Agency (US EPA), 2017, 2022b); however, the results of these methods have been curated by NIEHS as provided in the Integrated Chemical Environment (ICE) and thus present the most appropriate

Table 2. Uterotrophic method results for benzophenone and its metabolites by Nakagawa and Tayama (2001) and Yamasaki (2002)

Test Article	Subcutaneous Dose (mg/kg-day)	% Change Uterine wt
Control	0	0
4-HBP	40	+17[a]
4-HBP	100	+43*[b]
4-HBP	200	+51**[a]
4-HBP	200	+62*[b]
4-HBP	400	+91*[b]
4-HBP	800	+163**[a]
BP	400	+6[b]
BH	400	0[b]
EE	0.01	+294[a,b]
4-HBP	40 + EE	+3.1[a]
4-HBP	200 + EE	−21.1*[a]
4-HBP	800 + EE	−11.7[a]
4-HBP	TMX + EE	−24.2**[a]

[a] Relative wet weight: 4-HBP, 4-hydroxybenzophenone; BP, benzophenone; BH, benzhydrol; EE, ethinylestradiol; TMX, tamoxifen.
[b] Absolute wet weight.
* $P < 0.05$.
** $P < 0.01$.

location to access these data (United States Environmental Protection Agency (US EPA), 2022b). Please see Supplementary Table S-5 for the full list of available methods and Table 3 for assay results as extracted from the ICE v 2.0.

Is guidance available for the method(s)? (3.1)

Yes. The ToxCast suite of endocrine methods has been well described by Judson et al. (2015) and Browne et al. (2015); however, specific details for interpretation of the method results were limited to the method documentation as summarized by United States Environmental Protection Agency (US EPA) (2022c). Importantly, this battery of assays has not been validated to replace Level V multigeneration assays for endocrine evaluation but have been validated for replacement of the Utertrophic (OECD TG 440) screening assay that evaluates estrogen agonism. The basis for inclusion in this case study, was to determine whether the RfD derived from traditional animal assays would be protective of potential endocrine effects based on data from *in vitro* assays investigating estrogen agonism.

Are metabolite data available? (3.5)

Yes. ToxCast data (Tox21 methods) for the endocrine active metabolite, *p*-hydroxybenzophenone, were available. See answer to decision point 1.3 for this case study for a discussion on BP metabolism and relative endocrine activity.

Was the method conducted in a dose-response format? (3.6)

Yes. The ToxCast methods were conducted in a dose-response format.

Is external exposure known? (4.1)

No. Although monitoring data for drinking water extractives for BP are available, human exposure in the field in the use of certified products is not specifically known. As such, an RfD approach is applied.

Perform reverse dosimetry (IVIVE) to find human equivalent administered dose (EAD) (4.2)

RStudio (v. 3.6.3) was used to estimate the human EAD for the 1C and Solve 3comp physiologically based kinetic models from the NTP/NIEHS Integrated Chemical Environment (ICE). The models are further described in brief in the NTP ICE user guide (National Toxicology Program (NTP), 2021). To run the models, the fraction of chemical unbound to plasma protein (f_u), the intrinsic clearance from the liver (Cl_{int}), and the molecular weight are required (see Table 4). Casey et al. (2018) demonstrated that the application of the f_u is an equally effective way to establish free available test compound that accounts for binding of test compound to proteins and the method plate in comparison to the application of an enrichment factor (EF) by Armitage et al. (2014). The three-compartment model (Solve_3comp) estimates the maximum concentration (C_{max}) after acute and subchronic exposures and is influenced by repeated exposure; therefore, a 24-hour interval was used in the model. The duration of exposure used to estimate the C_{max} was 90 days as a comparative result to the available subchronic animal data.

In this QIVIVE assessment, the activity concentration at cutoff (ACC) was conservatively utilized as the point of departure for derivation of an EAD as this represents method activity above the method noise; however, in standard practice, the AC50 (activity concentration at 50% activity) is more routinely utilized as has been correlated to lowest effect levels (LELs) for *in vivo* data (Rotroff et al., 2010). However, for identification of a point of departure for risk assessment, it is not certain whether the ACC is well correlated to what would be considered a NOAEL or benchmark dose (BMD) that is the preferred point of departure to derive a human RfD. Further, the threshold of adversity for estrogen receptor ligand binding observed *in vitro* is not well defined.

Consider uncertainty: population and method variability (4.4)

Standard uncertainty factors according to NSF/ANSI/CAN 600 risk assessment procedures (consistent with US EPA) include interspecies, intraspecies, LOAEL to NOAEL (as applicable), subchronic to chronic, and database uncertainty. A standardized approach for application of uncertainty factors in the use of *in vitro* data to derive reference doses has not been identified. As the cell lines for the ToxCast endocrine methods are majority human, application of an interspecies uncertainty factor was deemed not applicable. A 10× factor for intraspecies uncertainty was maintained given potential variability in human response as compared to a homologous cell line. As the ACC was applied, which is proposed as a point of departure for any receptor activity, a LOAEL to NOAEL uncertainty factor was deemed not appropriate. Given the QIVIVE was run under a 90-day format, a subchronic to chronic uncertainty factor was not deemed applicable where a chronic animal study of ≥90 days is the acceptable study duration. As the nature of the endpoint is specific to estrogen agonism and the estrogen receptor battery has been validated as a replacement to *in vivo* estrogen receptor screening methods (Browne et al., 2015), a database uncertainty factor was also deemed not applicable. However, uncertainty associated with method variability was considered relevant to the assessment. Considering data from Sipes et al. (2017), comparison of the maximum human plasma concentration (C_{max}) of predicted *in silico* values using the httk package to measured human *in vivo* values gathered in DrugMatrix for 491 Tox21 chemicals and 613 dosing scenarios at pharmacologically relevant doses, found that

Table 3. *In vitro* endocrine method results and comparative reference doses (RfD) for benzophenone and 4-hydroxybenzophenone

Method Name (Exposure Duration)	Activity (ACC, μM)		Estimated Human RfD (mg/kg-day)[c]			
	Benzophenone[a]	4-Hydroxy-benzophenone[b]	Benzophenone		4-Hydroxybenzo-phenone	
			1C	HT3C	1C	HT3C
OT_ER_ERaERa_0480 (8 h)	53.985	Not tested	2.936	0.063	—	—
OT_ER_ERaERa_1440 (24 h)	Inactive	Not tested	—	—	—	—
OT_ER_ERaERb_0480 (8 h)	43.252	Not tested	2.353	0.050	—	—
OT_ER_ERaERb_1440 (24 h)	71.542	Not tested	3.891	0.083	—	—
OT_ER_ERbERb_0480 (8 h)	39.407	Not tested	2.143	0.046	—	—
OT_ER_ERbERb_1440 (24 h)	Inactive	Not tested	—	—	—	—
OT_ERa_EREGFP_0120 (2 h)	Inactive	Not tested	—	—	—	—
OT_ERa_EREGFP_0480 (8 h)	Inactive	Not tested	—	—	—	—
ATG_ERa_TRANS_up (24 h)	12.969	Not tested	—	—	—	—
ATG_ERE_CIS_up (24 h)	27.827	Not tested	1.514	0.032	—	—
Tox21_ERa_BLA_Agonist_ratio (16 h)	Inactive	30.594	—	—	1.437	0.026
Tox21_ERa_BLA_Antagonist_ratio (8 h)	Inactive	72.777	—	—	3.418	0.062
Tox21_ERa_LUC_BG1_Agonist (24 h)	Inactive	2.656	—	—	0.125	0.002
Tox21_ERa_LUC_BG1_Antagonist (48 h)	Inactive	70.996	—	—	3.334	0.060
TOX21_AR_BLA_Antagonist_ratio	Inactive	51.2927	—	—	2.409	0.043
TOX21_AR_LUC_MDAKB2_Antagonist	Inactive	30.0462	—	—	1.411	0.025

[a] Cytotoxicity limit = 8.04 μM.
[b] Cytotoxicity limit = 1000 μM.
[c] Derived from the Equivalent Administered Dose (EAD) estimated using the 1C and HT3C QIVIVE PBK models from ICE v 2.0 from the Activity Concentration at Cut-off (ACC) of the *in vitro* methods. A 100× total uncertainty factor [10× method uncertainty (Sipes, 2017) and 10× intraindividual variability] was applied.

Table 4. Quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) inputs for benzophenone and 4-hydroxybenzophenone

Chemical	CASRN	LogP	Cl_{int}[a]	fu[b]	MW[c]
Benzophenone	119-61-9	3.18	0.9298	0.0654	182.073
p-Hydroxybenzophenone	1137-42-4	3.07	0.796	0.033	198.068

[a] Intrinsic clearance; calculated as the log of observed clearance (μL/min/10^6 cells) from human *in vitro* metabolic methods (substrate depletion method) for p-t-butylphenol and benzophenone (Wetmore et al., 2015); predicted by QSAR model for 4-hydroxybenzophenone.
[b] Fraction unbound measured using the rapid equilibrium dialysis method (Wetmore et al., 2015); f_u was used per Casey et al. (2018).
[c] Molecular weight (g/mol).

80% of the *in silico* derived C_{max} values were within 10× of the observed values and 12% were overpredicted suggesting a significant breadth are included within an order of magnitude. Provided that measured values for physicochemical parameters were applied for BP and *p*-hydroxybenzophenone, a 10× uncertainty factor for method variability (as it relates to predicted values) was applied.

Estimate reference dose (4.5)

See Table 6 for comparative reference doses that were calculated from the human equivalent administered doses divided by an uncertainty factor of 100× (10× for interindividual; 10× for method variability per Sipes, 2017). The values in all cases, except one (the Tox21_ERa_LUC_BG1_Agonist (24 h) assay for p-hydroxybenzophenone that identified an RfDof 0.002 mg/kg-day), exceeded the comparative reference dose of 0.02 mg/kg-day derived from animal studies. Remarkably, the 3-compartment model provided reference doses of 0.032–0.083 mg/kg-day for BP and 0.002–0.062 for *p*-hydroxybenzophenone providing reference doses within an order of magnitude of the comparative animal RfD. Although individual method sensitivity and concordance of activity concentrations to adversity of apical endpoints in *in vivo* studies has not been well characterized, quantification of the EADs from these methods provides an indication of biological activity after exposure and increases confidence that the derived RfD is protective of public health.

Areas of uncertainty in the assessment

Challenges in data interpretation included understanding the comparative sensitivity across assays that was confounded by reported values as normalized to the positive control (eg, Odyssey Thera, OT, methods are normalized to 17β-estradiol and 4-hydroxytamoxifen) or as compared to the vehicle controls (Attagene, AT, methods). Further, the methods evaluate multiple estrogen receptor dimers (α, β, αβ) that may have varying affinities and differential expression in the human body that is further complicated by a lack of differentiation of agonist vs antagonist activity in some assays. As previously discussed, a standardized approach for the application of uncertainty factors in the context of *in vitro* methods has not been determined especially considering that methods are specific to limited biological pathways and utilize a homogenous cell culture, and thus the uncertainty factors as applied in this assessment are subject to refinement. Lastly, although the PBK models applied for the reverse dosimetry used measured data for the unbound fraction and metabolism of BP, the models have not been validated to observed human toxicokinetic data for BP or *p*-hydroxybenzophenone. Therefore, it is uncertain whether the one-compartment model or the 3-compartment model is a better representation of plasma concentration in humans. Overall, the QIVIVE assessment provides a conservative estimate of comparative human EAD and RfDs based on estrogen receptor activity (whether it be antagonistic or agonistic) that suggests a threshold of activity. Importantly, the ToxCast suite of endocrine assays has not been validated to replace definitive animal assays, but does add to the WoE, especially for MoA assessment, for the risk assessment of BP.

Case study 2: N-methylmorpholine N-oxide (NMMO)

Problem formulation (1.1)

An *in vitro* approach was developed to evaluate the reproductive toxicity of an industrial solvent, N-methylmorpholine N-oxide

Table 5. Dow cheminformatic predictions for XU-18840.00

Model Endpoint/Activity	Potential Interaction	Negative for Interaction
Reactivity		√
Chelation		√
Surfactant		√
Anticoagulant		√
Skin Sensitization	√[a]	
Transient receptor potential cation channel V1 (TRPV1)		√
Aconitase inhibitor		√
Mitochondria complex I, II, III, IV, or V		√
Mitochondrial protonophore	√[b]	
Cyanide (CN) or hydrogen sulfide (H_2S) releaser		√
Cardiac glycoside		√
Aryl hydrocarbon receptor (AhR)		√
Estrogen receptor (ER)		√
Androgen receptor (AR)		
Aromatase		√
Pyrethroid		√
Acetylcholine receptor (AChR) (muscarinic or nicotinic)		√
Acetylcholinesterase (AChE)	√[c]	
γ-Aminobutyric acid receptor (GABA)		√
Serotonin receptor		√
Glycinergic receptor		√

Dow cheminformatic models are developed for good balance accuracy, but favor sensitivity over specificity to minimize potential "false negative" findings.
[a] Skin sensitization identified as a possible target based on the formation of a potentially reactive metabolite; however, this metabolite was not identified with in *in vitro* metabolism studies using human skin S9 or human liver microsomes.
[b] If mitochondrial protonophore interaction occurs, this was judged to be at high concentrations by Dow's subject matter expert in cheminformatics. This endpoint was assessed *in vitro* using the Cell Stress Panel (Cyprotex, Watertown, Massachusetts).
[c] Acetylcholinesterase interaction was deemed unlikely by Dow's subject matter expert in cheminformatics. This endpoint was assessed in vitro with the Safety47 battery (Eurofins, San Diego, California).

Table 6. *In vitro* assay results for XU-18840.00

Endpoint (*In vitro* Assay)[a]	Negative
Skin irritation (EpiDerm Assay)	√
Genotoxicity (Ames)	√
Genotoxicity (Rat lymphocyte chromosomal aberration test; RLCAT)	√
Genotoxicity (Hypoxanthine-guanine phosphoribosyltransferase in CHO cells HGPRT)	√
Eye irritation/Corrosion (bovine corneal opacity and permeability [BCOP])	√
Phototoxicity/Cytotoxicity (3T3 neutral red uptake [NRU])	√
Dermal sensitization	√[b]
Safety47 battery	√[c]
Cell stress panel	√[d]
High throughput transcriptomics	√

[a] Tested concentrations vary by assay, but bioactive concentrations were ≥60 μM, which is considerably greater than 100× estimated internal doses.
[b] Negative for skin sensitization based on 2 *in vivo* studies conducted prior to 2013.
[c] Assay results did not indicate any bioactivity concerns, including no interaction with acetylcholinesterase. The Safety47 battery of assays is shown in Supplementary Table S-6.
[d] Small decreases in extracellular acidification rate (ECAR; a measure of mitochondrial stress) at the highest concentration only; all other assay responses were negative.

(NMMO) (Clewell *et al.*, 2022). NMMO is an industrial solvent, shown to cause decreased sperm production and reduce fertility in rats at doses > 50 mg/kg/day. Regulatory concerns center on whether these effects are relevant to the human, and whether the human is likely to be more, less or similarly sensitive to NMMD than the rat. Thus, the 2 questions are:

- Is NMMO likely to be a spermatotoxin in humans?
- If NMMO is a spermatotoxin in humans, are humans or rats more sensitive to it?

Key components (1.2)

Key components of the assessment are comparing rat and human 1) pharmacodynamics and 2) pharmacokinetics, as either component could lead to differences in species sensitivity. Of interest, but not as vital to the purpose of the assessment, is any information on MoA.

Consider toxicokinetics (1.3)

Very little information was available on the NMMO disposition and metabolism due to a lack of information on *in vivo* or *in vitro* metabolism or blood profiles. For this reason, an evaluation of parent chemical clearance in rat and human hepatocytes was performed. In addition, there was some concern that NMMO may degrade in the cell culture media. As a result, a stability study was performed in which NMMO concentration in the cell culture medium of the cell cultures was measured at various culture days. NMMO concentration in the media was stable over the 21-day experiment.

Is the adverse outcome known? (2.1)

Yes. The adverse outcome was reduced mature sperm number in sexually mature male rats.

Is the AOP/MoA known? (2.2)

The MoA is not fully defined for NMMO. However, 2 mechanisms can be proposed for the reduced sperm number: (1) spermatotoxicity (sperm cell death) or (2) disruption of germ cell differentiation. Thus, we targeted *in vitro* systems that could measure these endpoints. Since spermatocyte toxicity may be achieved by direct toxicity to the differentiating germ cell itself, or through disruption of the somatic environment, *in vitro* systems were sought that could evaluate both possibilities.

Identify KEs to measure (2.3)

While a formal AOP was not developed for this work, the spermatogenesis cycle has been well-described across species (eg, Cheng and Mruk, 2010) and KEs for spermatotoxicity could be identified from the published literature, including:

- Direct spermatotoxicity: germ cell number (live/dead cells)
- Direct inhibition of spermatocyte differentiation: markers of germ cell differentiation (RNA)
- Indirect inhibition of spermatocyte differentiation/viability: markers of testes blood barrier, somatic cell health

Is there an appropriate method available? (2.4)

Two methods were identified. The first measures human germ cell differentiation from committed germ cell to haploid spermatid in an isolated human germ cell line (Easley et al., 2012) (University of Georgia, Athens, Georgia). This model includes germ cell lines generated from multiple genetic backgrounds, allowing evaluation of interindividual susceptibility). The second method measures germ cell differentiation in a 3-dimensional seminiferous tubule *ex vivo* model in rat, monkey, and human

(BioAlter®, Kallistem, France). This model tests markers of germ cell, somatic cell and testes: blood barrier health over a period of 3 weeks. Because these studies were performed during the initial stages of the Covid 19 pandemic and all voluntary surgical procedures were canceled, human tissues were not available for this study. Thus, the non-human primate model served as a surrogate for human response, since the monkey has been shown to be a better model of the human spermatogenesis process, quantitatively and qualitatively, than the rodent (Fayomi and Orwig, 2018). Both systems measure the differentiation process from committed germ cell to haploid spermatid, which is the entire cycle that occurs in the testes. Models are not available for the spermiogenesis process that occurs in the epididymis, wherein the round spermatids become mature, mobile sperm. However, the delayed timing of the effects on sperm number indicated that the spermatotoxic effects of NMMO were occurring early in the spermatogenesis cycle, and thus, the existing models would cover the time-period of interest for this study (Clewell *et al.*, 2022).

Are NAMs data available? (2.5)

No.

Generate data (2.7)

Dose response and time-response data were generated for NMMO. Methylmethacrylate (MMA), a known reproductive toxin in rodent models, was used as a positive control to demonstrate method utility.

Method interpretation

Is guidance available for the method? (3.1)

Yes. These methods were described in the published literature and detailed standard operating procedures.

Is the applied method known to be scientifically robust? (3.2)

Yes.

Is there a role for bioactivation? (3.4)

No. Parent chemical clearance studies demonstrated that NMMO is not metabolized.

Reverse dosimetry (4.2)

Metabolism studies were performed by modeling parent chemical clearance in human and rat hepatocytes to support IVIVE calculations (Pharmacelsus, Germany). No metabolism was observed in either species. Read-across was used to parameterize the urinary excretion and blood protein binding parameters. IVIVE was performed (Yoon et al., 2015) for both the human and the rat.

Estimate reference dose (4.5)

Cell viability, germ cell differentiation (RNA, flow cytometry) and somatic environment (RNA) data were collected in the rat and nonhuman primate 3D seminiferous tubule *ex vivo* model (Kallistem) and cytotoxicity, oxidative stress, cell cycle and viability were measured in the human monoculture germ cell differentiation model. Testes: blood barrier integrity (TEER) was measured in the rat seminiferous tubule model. Both models follow spermatogenesis from committed germ cell to haploid spermatid. In the rat 3D seminiferous tubule model, effects were seen on later stage spermatocytes, somatic cells and gap junctions, as well as in the measures of membrane barrier integrity. These *in vitro* effects were used to define a PoD and were coupled with IVIVE to derive an oral equivalent dose (OED) (Rotroff et al., 2010; Yoon et al., 2015). Human OEDs were calculated from the monkey *in vitro* PoD using human body weight and metabolism, assuming that pharmacodynamic response would be similar in the monkey and human (Fayomi and Orwig, 2018).

Oral equivalent doses (OED) were calculated according to Rotroff et al. (2010) and were developed for rats, humans and monkeys in order to compare species. An RfD was not calculated in this example; however, the OED would be considered a derived PoD from which an RfD could be calculated after applying appropriate uncertainty factors. Neither species showed evidence of direct cytotoxicity at any dose of NMMO. However, the rat model showed dose-dependent decreases in secondary spermatocyte populations at OEDs $\geq$ 89 mg/kg/day, as well as reduced expression of mRNA markers for several stages of spermatogenesis (spermatogonia, pachytene spermatocytes, round spermatids) at an OED of 267 mg/kg/day NMMO. In contrast, the monkey model did not show dose-dependent decreases in these same RNAs at OEDs up to 1376 mg/kg/day. Indeed, the monkey consistently demonstrated increased expression of RNAs that were decreased in the rat. The opposite trends in the functional markers of the spermatocyte populations indicate that NMMO is unlikely to have similar effects on rat and monkey sperm count. Moreover, mRNA expression for Sertoli cells and tight junctions, were mildly increased with NMMO treatment in the monkey, but mildly decreased in the rat. Given that there are significant species differences between rat and human spermatogenesis, and monkey spermatogenesis is more similar to the human (Fayomi and Orwig, 2018), these qualitative differences in germ cell and somatic response indicate that human response is unlikely to be similar to the rat. Further, as treatments with OEDs higher than the *in vivo* dose limit (1376 mg/kg/day) did not show reduced sperm cell populations in the monkey, we would conclude that NMMO is unlikely to have anti-spermatogenic effects in the human.

Case study 3: XU-18840.00 - Development of a new cosmetic ingredient

Problem formulation (1.1)

In the EU, it is prohibited to use animal test data generated after 2013 to support the safety assessment of cosmetic ingredients (European Union (EU), 2009). In addition, the safety requirements for cosmetic ingredients are not strictly defined. However, the International Cooperation on Cosmetics Regulations (ICCR) has developed a guidance for Next-Generation Risk Assessments (NGRA) using NAMs for cosmetics (Berggren et al., 2017; Dent et al., 2018; International Cooperation on Cosmetic Regulation (ICCR), 2018, 2021). In 2020, Baltazar *et al.* applied the International Cooperation on Cosmetic Regulation (ICCR) (2018) concept by using an exposure-led, hypothesis-driven approach for a NGRA for a cosmetic ingredient case study (coumarin) used in face cream and body lotion. New cosmetic ingredients require a nontargeted assessment as the toxicity has not been characterized, examining a broader biological space with safety determinations based on a MoS relative to the lowest point of departure (PoD) concentration showing bioactivity in NAM assessments. Case study 3 describes a new cosmetic ingredient for face cream, XU-18840.00, which required a non-animal safety assessment to be sold in the EU; thus, the problem was to adapt the nontargeted NAM-based approach described by Baltazar to evaluate the safety of this ingredient. Ultimately, the decision was made not to

market this product in the EU and thus, animal data can be used as part of the safety assessment. This provides a case study in which safety determinations based on NAMs can be anchored in some animal data; thereby providing learning opportunities to support future assessments with alternative models.

Identify key components (1.2)

XU-18840.00 was designed as a cosmetic ingredient for use in face cream. Given direct consumer exposures to XU18840.00, it is important to examine bioactivity across a broad range of potential toxicological targets, including complex hazard endpoints like systemic toxicity. To do this, NAMs must examine a broad range of potential biological activity (while being practical) and results must be integrated in a WoE approach to assess XU-18840.00 safety. First, estimated exposure levels to XU-18840.00 are determined based on use patterns for face creams (Scientific Committee on Consumer Safety (SCCS), 2018), the concentration of XU-18840.00 in these products (0.1%; 1.4 μg/cm^2 skin), and the dermal penetration of XU-18840.00 in the relevant carrier matrix was determined using fresh human skin samples. Next, a suite of *in silico* and *in vitro* NAMs are used to identify potential bioactivity, generate multiple concentration-response curves and identify the lowest PoD concentration (see Tables 5 and 6). Internal dose was estimated based on face cream exposure and dermal penetration of XU-18840.00 (limited metabolism alleviated the need to consider metabolites). Forward dosimetry was used to compare the internal dose with the *in vitro* PoD to determine a MoS for XU-18840.00. This NextGen safety assessment can be supported with some animal data, which are available for this example.

Consider toxicokinetics (1.3)

Initially, a pharmacokinetic model for XU-18840.00 was developed using GastroPlus™ software. This model predicted that XU-18840.00 would reach steady state quickly after repeated dermal exposures and that metabolism of XU-18840.00 was limited. Subsequent *in vitro* experiments incubating XU-18840.00 with human liver microsomes and human skin S9 confirmed that there were no reactive metabolites formed, but rather only one minor glucuronide metabolite. This allowed the bioactivity evaluations to focus on the parent compound (XU-18840.00), thereby minimizing the impact of poor metabolic capacity in various *in vitro* methods. The lack of toxic metabolite(s) is indirectly supported by the results of an *in vivo* OECD TG 408-compliant 90-day oral study in rats that were treated with approximate dose levels of 0, 62, 186 and 624 mg/kg/day XU-18840.00. In this study, limited toxicity was observed at 624 mg/kg/day XU-18840.00 (further details are provided in 2.7 below).

Is the adverse outcome(s) (endpoint) known? (2.1)

No. For a NAM-based NextGen assessment, the potential bioactivity of XU-18840.00 was not known and thus, a nontargeted endpoint evaluation was conducted.

Are NAM data available? (2.5)

No. NAM data were not available as this was a new cosmetic ingredient.

Generate NAM data (2.7)

Computational (*in silico*) modeling was used initially to examine the potential bioactivity of XU-18840.00. Read-across compounds were judged to be inadequate and thus, XU-18840.00 was evaluated using Dow Cheminformatics profilers and models to predict toxicological hazards based on chemical structure. Dow's models are conservative, favoring sensitivity over specificity, to avoid false negative outcomes. The majority of endpoints examined were negative, but there were potential flags for skin sensitization (due to a predicted reactive metabolite, which was not formed when examined experimentally using *in vitro* microsomal incubations. These *in vitro* results were further supported by *in vivo* data, which included negative local lymph node and Buehler assays). Cheminformatic models also predicted potential mitochondrial interactions and acetylcholinesterase inhibition (Table 5) (Wijeyesakere et al., 2018, 2019, 2020).

Next, XU-18840.00 was examined in a series of *in vitro* methods to characterize potential bioactivity (Table 6). As outlined in Baltazar et al. (2020), these methods included *in vitro* assays for skin irritation (EpiDerm™), genotoxicity (Ames +/- metabolic activation; rat lymphocyte chromosomal aberration test (RLCAT); hypoxanthine-guanine phosphoribosyltransferase (HGPRT) gene mutation test in Chinese hamster ovary cells), eye corrosion (Bovine corneal opacity and permeability), and photo/cytotoxicity (3T3 neutral red uptake). These *in vitro* assays have associated OECD test guidelines and are accepted in place of animal data. A Safety47 battery (Eurofins, San Diego, California; see Supplementary Table S-6) was included to look for interactions with target proteins of concern (ie, often associated with adverse drug reactions); this battery included an examination of potential acetylcholinesterase interaction, which was negative at all concentrations tested (≤ 60 μM). The Safety47 battery did not identify any bioactivity concerns, although it did identify the intended bioactivity for which XU18840.00 was designed (ie, antioxidant response). A cell stress panel (Cyprotex, Watertown, Massachusetts; see Supplementary Table S-7) was also included, which examined nonspecific cell stress and cytotoxicity parameters including numerous endpoints to evaluate mitochondrial function. Only one parameter was altered in the cell stress panel, extracellular acidification rate (ECAR), which indicates a decrease in glycolysis at the highest concentration of XU-18840.00 (60 μM). This activity was used as the lowest PoD. The remaining *in vitro* evaluation, high throughput transcriptomics, is in progress.

The *in vivo* results support the general absence of effects in the NAM-based assessments. In addition, data are available from a 90-day oral toxicity study in which there were treatment-related decreases in body weight (≤7%) and feed consumption in female rats at the high-dose (~624 mg/kg/day). Increased liver weights were reported in high-dose males and females, and males also had very slight hepatocellular hypertrophy; these liver effects were considered adaptive as there were no additional histopathology findings. Thus, the high-dose was considered a no observed adverse effect level (NOAEL) for males and the mid-dose was the NOAEL for females. In addition, XU-18840.00 was non-sensitizing in both the local lymph node (LLNA) and Buehler assays. Both acute oral and dermal *in vivo* toxicity values exceeded the limit doses (5000 and 2000 mg/kg, respectively).

Is external exposure known? (4.1)

Yes. External exposure for face creams is well described by Scientific Committee on Consumer Safety (SCCS) (2018) and requires a determination of the amount of ingredient (mg) in contact with the skin per each occasion using variables such as frequency of use, ingredient use level, "leave on" versus "rinse off", and exposure duration per occasion. For XU-18840.00, the use level in face cream was set at 0.1%; thus, external exposure was calculated using the formulas in Scientific Committee on Consumer Safety (SCCS) (2018) with an estimate of 0.77 mg XU-

18840.00 in contact with the skin per use. Common use levels for an analogue antioxidant in skin cream is 0.5-1%.

Perform forward dosimetry (PBK) to find estimated internal dose (4.3)

Exposure data in mg was determined in step 4.1. For internal dose estimates, *in vitro* dermal penetration was determined to be ≤ 3.5% using radiolabeled XU-18840.00 as formulated in the commercial product and fresh split thickness skin from 2 human donors (Pharmaron UK Limited, Rushdon, Great Britain). Partial thickness skin samples were used to provide high-end estimates of dermal penetration. *In vitro* human metabolism studies with liver and skin S-9 identified XU-18840.00-glucuronide as the only measurable metabolite; no reactive metabolites were detected. These data were used to refine the pharmacokinetic model developed in GastroPlus™ (step 1.3) to estimate an internal dose of 9.4 nM. Comparatively, the same method was used to estimate the internal C_{max} in humans using the *in vitro* internal clearance rate data and after an oral animal exposure of 600 mg/kg-day. The internal dose from this oral exposure was estimated as 11.4–11.8 μM, which is 1255× higher than the estimated internal dose following dermal exposure, suggesting that the NAMs-based approach is conservative and protective.

Consider uncertainty: population and method variability (4.4)

Interspecies extrapolation is of lower concern for the NAM-based assessments as human-based assays/models were used to screen for most bioactivities, including those driving the PoD. The potential for active metabolites is not supported by available data. However, limitations on the biological endpoints evaluated for potential activity contributes to uncertainty. In addition, population distributions for the pharmacokinetic model represent one source of uncertainty. Monte Carlo analysis can be used to generate multiple MoS values using steady state blood distributions based on variable input parameters and PoD distributions from *in vitro* methods that used multiple sampling time points (see Baltazar *et al.*, 2020).

Available *in vivo* data support the results of the NextGen safety assessment. For example, the 90-day study was conducted orally as a 'worst case' scenario and to examine metabolites. *In vitro* dermal studies using fresh human skin estimate XU-18840.00 penetration at ≤ 3.5% of the administered dose; thus, the 90-day oral study is likely protective of internal doses achieved in humans with dermal applications.

Estimate RfD or calculate MoS (4.5)

The MoS is determined based on the ratio between the lowest PoD (step 2.7) and the internal dose (step 4.3). A large MoS indicates that the internal dose is markedly less than the bioactive PoD. An adequate MoS has not been defined; however, Paul Friedman et al. (2020) has reported that *in vitro* PoDs from 24 h exposures across a diverse battery of assays are generally as conservative as *in vivo* studies and can be used to determine a quantitative PoD similar to traditional animal repeat-dose and developmental/reproductive toxicity studies. Thus, an MoS equal to or greater than 100 has been proposed. At this time, the current PoD supports the use of XU-18840.00 at the proposed concentration of 0.1% in face cream. According to the NextGen Safety Assessment, if the transcriptomic PoD is lower (ie, XU-18840.00 has bioactivity at a lower concentration than currently identified), there are 3 potential outcomes: (1) the MoS remains sufficient (ie, ≥ 100×), supporting the use of XU-18840.00 at the current proposed level (0.1%); (2) the MoS is insufficient, such that the concentration of XU-18840.00 that can be used in face cream must be decreased; or (3) the MoS is insufficient such that XU-18840.00 cannot be used in this cosmetic application. Further, as the internal dose at the LOAEL from the 90-day study (ie, 11.4 μM) is > 1000× the internal dose for the estimated dermal exposure and is comparable to the NAMs PoD for bioactivity (ie, 60 μM), the oral data support the safety in use of XU-18840.00 at the proposed 0.1% level and the feasibility for use of NAMs data for this case study.

Conclusions

The framework that has been proposed within this document and the associated resources that have been identified are intended to foster the use of NAMs data in risk assessment in a standardized way for both novice and experienced risk assessment practitioners. The case reports that have been provided illustrate the usability and versatility of the framework. The specific *in vitro* systems that are used address questions of chemical mode-of-action, metabolism, and/or toxicokinetics aligned with the respective AOPs relevant to the endpoints of interest. Applying QIVIVE to extrapolate *in vitro* tissue exposures to human equivalent doses establishes that the complexities in extrapolation and data interpretation can be managed, demonstrating the utility of the proposed framework while also highlighting potential sources of human population bias and uncertainty. Further, these case reports demonstrate that incorporating *in vitro* assays, toxicogenomics, bioactivity, toxicokinetic data, exposure, and other biological method data within non-animal approaches is a viable and practical option to highlight human relevance and add significant value to a WoE evaluation in risk assessment. However, gaining confidence in NAMs is critical to their implementation and widespread regulatory acceptance. This article facilitates that process by sharing NAM case studies that demonstrate their utility based on an understanding of human biology.

For many toxicological endpoints, specific NAMs test guidelines, protocols, best practices, or overarching defined approaches are lacking for the use of non-animal methods in regulatory applications; however, general quality and methodological guidelines are available for development and to support interpretation of NAMs (eg, Organization for Economic Co-operation and Development (OECD) (2018b)). In addition, workshops that lay the foundation for future guidance documents (eg, the use of *in vitro* metabolic studies in pesticide risk assessment by European Food Safety Authority (EFSA) *et al.* (2019)) and some guidance documents including Organization for Economic Co-operation and Development (OECD) (2010) Test No. 417- "Toxicokinetics" identify "supplemental approaches", that highlight the use of *in vitro* methods to evaluate substance metabolism. Web-based tools for modeling toxicokinetic parameters are also now available (Bossier et al., 2020). More recently, the Organization for Economic Co-operation and Development (OECD) (2021d) has released guidance geared towards the regulatory community to increase confidence in the use of data derived from non-animal methods that can be incorporated into PBK models. These models, although not known to be explicit requirements in any regulatory jurisdiction, are routinely used to incorporate physiological or biochemical mechanistic parameters to predict internal exposure. Toxicokinetic data, from *in vitro* and *in silico* approaches, are now routinely submitted to regulatory agencies. Specifically, the US FDA has several guidance documents that include the use of *in vitro* or *in silico* approaches (United States Food and Drug Administration (US FDA), 2018, 2021a, 2020a,b). In

addition, a recent review by Stucki et al. (2022) provides further insight on the ability to use NAMs within US, EU and Canadian regulatory frameworks and a report of a recent workshop convened by the European Partnership for Alternative Approaches to Animal Testing (EPAA) identifies the key changes needed in regulatory frameworks, education, training, stakeholder engagement, and science to advance the use of NAMs in regulatory decision-making.

Over 350 000 chemicals and mixtures of chemicals are globally registered for production and use worldwide (Wang et al., 2020) that during their life cycle can be released into the environment, enter the food chain or otherwise contribute to human exposure. Food, in particular, has been shown to be a dominant source of exposure. Real-life exposure comprises the simultaneous or consecutive (co)-exposure to chemical substances originating from various sources, including pesticides, pharmaceuticals, personal care products and, most importantly, diet. This includes foodstuffs as well as environmental and food process contaminants, residues of pesticides, veterinary medicines and chemicals from food contact materials, food additives and others.

Existing legislation provides extensive information on the respective substances within their remit. However, it generally addresses the effects of single substances only, and this inherently carries the risk of missing potential mixture-induced effects or shifts in the dose-response curve associated with mixture exposures. Consequently, there is an appeal for the introduction of combined exposure in risk assessment and high throughput safety evaluations. The proposed NAM framework offers potential to use advances in science and technology for implementing but especially revolutionizing modern chemical risk assessment. For a given mixture, the question is whether combinatorial effects exist, if current regulatory measures are sufficiently protective, and whether new tools/approaches may be used to fill data gaps, enhance our understanding of potential mixture effects and improve public health protection (Tralau et al., 2021).

With advancing science and regulatory initiatives to move away from animal testing (United States Environmental Protection Agency (US EPA), 2021), increasing efforts are being made to develop non-animal based NAMs which mimic or resemble toxicity and disease pathways relevant to the humans. In the last decade, governments as well as different industries around the world have been investing significant amounts of resources in NAMs, which has led to significant progress in development and validation efforts (see Supplementary Table S-2). Despite the significant work, there remains in some cases, limited confidence in non-animal methods for PBK modeling and (Q)IVIVE, possibly due to low confidence in exposure data, internal dosimetry estimates and/or a desire to focus on hazard identification without consideration of exposure and toxicokinetics. As additional guidance and specific protocols with international acceptance become available, regulatory submissions utilizing NAMs approaches should continue to increase. Available resources, like this framework, that address how these tools can be used to support WoE evaluation and risk assessment are expanding. Coupled with understanding how to quantify and qualify uncertainty, we expect to see a growing reliance on nonanimal toxicokinetic assessments for regulatory application.

Future directions

Targeted safety evaluations applying NAMs within known or proposed AOPs and nontargeted safety assessments considering chemical bioactivity is demonstrated in this paper as a viable approach that significantly increases confidence in the WoE evaluation in risk assessment. However, full validation in use of NAMs in lieu of traditional animal testing and characterization of uncertainty (including predictivity in human populations vs individuals based on homogenous cell populations) remain ongoing areas of research. Although, proposals are made within the case studies presented in this document for how to manage uncertainties in the use of NAMs and QIVIVE, formal and comprehensive evaluation of the most appropriate uncertainty factors for *in vitro* assays and quantitative extrapolation to human equivalent doses have yet to be defined adequately. Despite the need to increase confidence in use of NAMs, NAMs approaches can actively contribute to the understanding of chemical exposures that can affect improvements in human health in the near term.

Supplementary data

Supplementary data are available at *Toxicological Sciences* online.

Acknowledgments

We are particularly thankful to Amalia Munoz and Gerard Bowe of the European Joint Research Centre for their editorial help with the figures and tables related to OECD GIVIMP.

Funding

No funding was obtained in the production of this manuscript by any of the authors.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

AltTox. (2019). *Methods, Approaches, Programs & Policies. Validation. AltTox.Org. 2019*. Available at: http://alttox.org/mapp/validation-2/. Accessed October 23, 2022.

Armitage, J. M., Sangion, A., Parmar, R., Looky, A. B., and Arnot, J. A. (2021). Update and evaluation of a high-throughput in vitro mass balance distribution model: IV-MBM EQP v2.0. *Toxics* **9**, 315.

Armitage, J. M., Wania, F., and Arnot, J. A. (2014). Application of mass balance models and the chemical activity concept to facilitate the use of in vitro toxicity data for risk assessment. *Environ. Sci. Technol.* **48**, 9770–9779.

Ball, N. *et al.* (2016). Toward good read-across practice (GRAP) guidance. *ALTEX* **33**, 149–166.

Bal-Price, A. (2018). Recommendation on test readiness criteria for new approach methods in toxicology: exemplified for developmental neurotoxicity. *ALTEX* 306–352. doi: 10.14573/altex.1712081.

Bal-Price, A., Crofton, K. M., Leist, M., Allen, S., Arand, M., Buetler, T., Delrue, N., FitzGerald, R. E., Hartung, T., Heinonen, T., *et al.* (2015). International STakeholder NETwork (ISTNET): creating a developmental neurotoxicity (DNT) testing road map for regulatory purposes. *Arch. Toxicol.* **89**, 269–287.

Baltazar, M. T., Cable, S., Carmichael, P. L., Cubberley, R., Cull, T., Delagrange, M., Dent, M. P., Hatherell, S., Houghton, J., Kukic, P., *et al.* (2020). A next-generation risk assessment case study for coumarin in cosmetic products. *Toxicol. Sci.* **176**, 236–252.

Barron, M. G., Lilavois, C. R., and Martin, T. M. (2015). MOAtox: A comprehensive mode of action and acute aquatic toxicity database for predictive model development. *Aquat. Toxicol.* **161**, 102–107.

Barton-Maclaren, T. S., Wade, M., Basu, N., Bayen, S., Grundy, J., Marlatt, V., Moore, R., Parent, L., Parrott, J., Grigorova, P., *et al.* (2022). Innovation in regulatory approaches for endocrine disrupting chemicals: The journey to risk assessment modernization in Canada. *Environ. Res.* **204**, 112225.

Bas, A., Burns, N., Gulotta, A., Junker, J., Drasler, B., Lehner, R., Aicher, L., Constant, S., Petri-Fink, A., and Rothen-Rutishauser, B. (2021). Understanding the development, standardization, and validation process of alternative in vitro test methods for regulatory approval from a researcher perspective. *Small* **17**, 2006027.

Bell, S. M., Chang, X., Wambaugh, J. F., Allen, D. G., Bartels, M., Brouwer, K. L. R., Casey, W. M., Choksi, N., Ferguson, S. S., Fraczkiewicz, G., *et al.* (2018). In vitro to in vivo extrapolation for high throughput prioritization and decision making. *Toxicol. In Vitro* **47**, 213–227.

Berggren, E., White, A., Ouedraogo, G., Paini, A., Richarz, A.-N., Bois, F. Y., Exner, T., Leite, S., van Grunsven, L. A., Worth, A., *et al.* (2017). Ab initio chemical safety assessment: a workflow based on exposure considerations and non-animal methods. *Comput. Toxicol.* **4**, 31–44.

Birch, H., Kramer, N. I., and Mayer, P. (2019). Time-resolved freely dissolved concentrations of semivolatile and hydrophobic test chemicals in in vitro assays—measuring high losses and crossover by headspace solid-phase microextraction. *Chem. Res. Toxicol.* **32**, 1780–1790.

Blaauboer, B. J. (2010). Biokinetic modeling and in vitro – in vivo extrapolations. *J. Toxicol. Environ. Health. B Crit. Rev.* **13**, 242–252.

Bossier, H., Chau, J., Ndour, C., Varewyck, M., Verbeke, T., and Vergucht, S. (2020). A web-based open source tool for toxicokinetic and toxicodynamic modelling. *EFS3* **17**, doi: 10.2903/sp.efsa.2020.EN-1926.

Breen, M., Ring, C. L., Kreutz, A., Goldsmith, M.-R., and Wambaugh, J. F. (2021). High-throughput PBTK models for in vitro to in vivo extrapolation. *Expert Opin. Drug Metab. Toxicol.* **17**, 903–921.

Bronzwaer, S., Catchpole, M., de Coen, W., Dingwall, Z., Fabbri, K., Foltz, C., Ganzleben, C., van Gorcom, R., Humphreys, A., Jokelainen, P., *et al.* (2022). One health collaboration with and among EU agencies – Bridging research and policy. *One Health* **15**, 100464.

Browne, P., Judson, R. S., Casey, W. M., Kleinstreuer, N. C., and Thomas, R. S. (2015). Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ. Sci. Technol.* **49**, 8804–8814.

Browne, P., Noyes, P. D., Casey, W. M., and Dix, D. J. (2017). Application of adverse outcome pathways to U.S. EPA's endocrine disruptor screening program. *Environ. Health Perspect.* **125**, 096001.

Casey, W. M., Chang, X., Allen, D. G., Ceger, P. C., Choksi, N. Y., Hsieh, J.-H., Wetmore, B. A., Ferguson, S. S., DeVito, M. J., Sprankle, C. S., *et al.* (2018). Evaluation and optimization of pharmacokinetic models for in vitro to in vivo extrapolation of estrogenic activity for environmental chemicals. *Environ. Health Perspect.* **126**, 97001.

Center for Disease Control (CDC). (2022). *One Health | CDC*. April 19, 2022. Available at: https://www.cdc.gov/onehealth/index.html. Accessed October 23, 2022.

Chang, X., Tan, Y.-M., Allen, D. G., Bell, S., Brown, P. C., Browning, L., Ceger, P., Gearhart, J., Hakkinen, P. J., Kabadi, S. V., *et al.* (2022). IVIVE: Facilitating the use of in vitro toxicity data in risk assessment and decision making. *Toxics* **10**, 232.

Cheng, C. Y., and Mruk, D. D. (2010). The biology of spermatogenesis: The past, present and future. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **365**, 1459–1463.

Clerbaux, L.-A., Albertini, M. C., Amigó, N., Beronius, A., Bezemer, G. F. G., Coecke, S., Daskalopoulos, E. P., del Giudice, G., Greco, D., Grenga, L., *et al.* (2022). Factors modulating COVID-19: A mechanistic understanding based on the adverse outcome pathway framework. *JCM* **11**, 4464.

Clewell, R. A., Leonard, J. A., Nicolas, C. I., Campbell, J. L., Yoon, M., Efremenko, A. Y., McMullen, P. D., Andersen, M. E., Clewell, H. J., Phillips, K. A., *et al.* (2020). Application of a combined aggregate exposure pathway and adverse outcome pathway (AEP-AOP) approach to inform a cumulative risk assessment: A case study with phthalates. *Toxicol. In Vitro* **66**, 104855.

Clewell, R. A., Clewell, H. J., Linakis, M. W., Easley, C. A., Langmo, J. N., Salley, J., Gentry, R., and Rücker, T. (2022). An in vitro approach to determine the human relevance of antispermatogenic effects of 4-methylmorpholine 4-oxide, monohydrate (NMMO) in rat reproductive toxicity studies. *Toxicol. In Vitro* **82**, 105365.

Coecke, S., Ahr, H., Blaauboer, B. J., Bremer, S., Casati, S., Castell, J., Combes, R., Corvi, R., Crespi, C. L., Cunningham, M. L., *et al.* (2006). Metabolism: A bottleneck in in vitro toxicological test development: The report and recommendations of ECVAM workshop 54. *Altern. Lab. Anim.* **34**, 49–84.

Coecke, S., Bernasconi, C., Bowe, G., Bostroem, A.-C., Burton, J., Cole, T., Fortaner, S., *et al.* (2016). Practical aspects of designing and conducting validation studies involving multi-study trials. In *Validation of Alternative Methods for Toxicity Testing* (C. Eskes and M. Whelan, Eds), Vol. **856,** pp. 133–163. Advances in Experimental Medicine and Biology. Springer International Publishing, Cham.

Cohen, S. M., Boobis, A. R., Dellarco, V. L., Doe, J. E., Fenner-Crisp, P. A., Moretto, A., Pastoor, T. P., Schoeny, R. S., Seed, J. G., and Wolf, D. C. (2019). Chemical carcinogenicity revisited 3: Risk assessment of carcinogenic potential based on the current state of knowledge of carcinogenesis in humans. *Regul. Toxicol. Pharmacol.* **103**, 100–105.

Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., Wiegers, J., Wiegers, T. C., and Mattingly,C. J. (2020). The comparative toxicogenomics database: Update 2021. *Nucleic Acids Res.* http://ctdbase.org/about/publications/#ctdpubs.

Deconinck, K. (2021). Concentration and market power in the food chain. In *OECD Food, Agriculture and Fisheries Papers*, No. 151. OECD Publishing, Paris. doi: 10.1787/3151e4ca-en.

Deisenroth, C., DeGroot, D. E., Zurlinden, T., Eicher, A., McCord, J., Lee, M.-Y., Carmichael, P., and Thomas, R. S. (2020). The alginate immobilization of metabolic enzymes platform retrofits an estrogen receptor transactivation assay with metabolic competence. *Toxicol. Sci.* **178**, 281–301.

De Jong, W. H., Carraway, J. W., Liu, C., Fan, C., Liu, J., Turley, A. P., Rollins, T. S., and Coleman, K. P. (2020). The suitability of reconstructed human epidermis models for medical device irritation assessment: A comparison of in vitro and in vivo testing results. *Toxicol. In Vitro* **69**, 104995.

de Jongh, E. J., Harper, S. L., Yamamoto, S. S., Wright, C. J., Wilkinson, C. W., Ghosh, S., and Otto, S. J. G. (2022). One health, one hive: A scoping review of honey bees, climate change, pollutants, and antimicrobial resistance. *PLoS One* **17**, e0242393.

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

DG SANTE. (2022). *Authorisation of Additives*. 2022. Available at: https://webgate.ec.europa.eu/foods_system/main/?sector=FAD&auth=SANCAS. Accessed October 23, 2022.

Doak, S. H., Jenkins, G. J. S., Johnson, G. E., Quick, E., Parry, E. M., and Parry, J. M. (2007). Mechanistic influences for mutation induction curves after exposure to DNA-reactive carcinogens. *Cancer Res.* **67**, 3904–3911.

Dourson, M., Ewart, L., Fitzpatrick, S. C., Barros, S. B. M., Mahadevan, B., and Hayes, A. W. (2022). The future of uncertainty factors with in vitro studies using human cells. *Toxicol. Sci.* **186**, 12–17.

Easley, C. A., Phillips, B. T., McGuire, M. M., Barringer, J. M., Valli, H., Hermann, B. P., Simerly, C. R., Rajkovic, A., Miki, T., Orwig, K. E., *et al.* (2012). Direct differentiation of human pluripotent stem cells into haploid spermatogenic cells. *Cell Rep.* **2**, 440–446.

European Commission (EC). (2019). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. The European Green Deal*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2019%3A640%3AFIN.

European Commission (EC). (2020). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Chemicals Strategy for Sustainability Towards a Toxic-Free Environment*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A667%3AFIN.

European Centre for the Validation of Alternative Methods (ECVAM). (2019). *EURL ECVAM Dataset on Alternative Methods to Animal Experimentation (DB-ALM)*. Available at: http://data.europa.eu/89h/b7597ada-148d-4560-9079-ab0a5539cad3.

European Centre for the Validation of Alternative Methods (ECVAM). (2022). *Home | EURL ECVAM - TSAR*. 2022. Available at: https://tsar.jrc.ec.europa.eu/.

European Food Safety Authority (EFSA), Hart, A., Maxim, L., Siegrist, M., Von Goetz, N., da Cruz, C., Merten, C., Mosbach-Schulz, O., Lahaniatis, M., Smith, A., *et al.* (2019). Guidance on communication of uncertainty in scientific assessments. *EFSA J.* **17**. doi:10.2903/j.efsa.2019.5520.

Elmeliegy, M., Udata, C., Liao, K., and Yin, D. (2021). Considerations on the calculation of the human equivalent dose from toxicology studies for biologic anticancer agents. *Clin. Pharmacokinet.* **60**, 563–567.

European Union (EU). (2009). *Regulation (EC) No 1223/2009 if the European Parliament and of the Council of 30 November 2009 on Cosmetic Products*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009R1223-20160812&from=EN.

European Union (EU). (2010). *Regulation (EU) No 911/2010 of the European Parliament and of the Council of 22 September 2010 on the European Earth Monitoring Programme (GMES) and Its Initial Operations (2011 to 2013), October*. doi:10.3000/17252555.L_2010.276.eng.

European Union (EU). (2022). *Nuno Lesson. Education and Training Platform for Laboratory Animal Science*. 2022. Available at: https://etplas.eu/lessons/nuno-lesson/.

European Union (EU) ToxRisk. (2022). *EU-ToxRisk - Test Methods*. 2022. Available at: https://www.eu-toxrisk.eu/page/media_items/test-methods8.php.

Fayomi, A. P., and Orwig, K. E. (2018). Spermatogonial stem cells and spermatogenesis in mice, monkeys and men. *Stem Cell Res.* **29**, 207–214.

Felter, S. P., Bhat, V. S., Botham, P. A., Bussard, D. A., Casey, W., Hayes, A. W., Hilton, G. M., Magurany, K. A., Sauer, U. G., and Ohanian, E. V. (2022). Assessing chemical carcinogenicity: hazard identification, classification, and risk assessment. Insight from a toxicology forum state-of-the-science workshop. *Crit. Rev. Toxicol.* **51**, 653–694.

Firman, J. W., Punt, A., Cronin, M. T. D., Boobis, A. R., Wilks, M. F., Hepburn, P. A., Thiel, A., and Fussell, K. C. (2021). Exploring the potential of ToxCast data in supporting read-across for evaluation of food chemical safety. *Chem. Res. Toxicol.* **34**, 300–312.

Gannon, A. M., Moreau, M., Farmahin, R., Thomas, R. S., Barton-Maclaren, T. S., Nong, A., Curran, I., and Yauk, C. L. (2019). Hexabromocyclododecane (HBCD): A case study applying tiered testing for human health risk assessment. *Food Chem. Toxicol.* **131**, 110581.

Gonzalez, E., Shah, P., Zakharov, A., Nguyen, D.-T., Torimoto-Katori, N., Sakamuru, S., Xia, M., Zhao, T., Obach, R. S., Hop, C., *et al.* (2018). Utilization of artificial intelligence to develop a QSAR model for predicting P450 mediated metabolic stability. *FASEB J.* **32**, 529.8–529.8.

Good Cell Culture Practice (GCCP). (2018). *In Guidance Document on Good In Vitro Method Practices (GIVIMP), by OECD, 161–161. OECD Series on Testing and Assessment*. OECD. doi:10.1787/9789264304796-16-en.

Halappanavar, S., van den Brule, S., Nymark, P., Gaté, L., Seidel, C., Valentino, S., Zhernovkov, V., Høgh Danielsen, P., De Vizcaya, A., Wolff, H., *et al.* (2020). Adverse outcome pathways as a tool for the design of testing strategies to support the safety assessment of emerging advanced materials at the nanoscale. *Part. Fibre Toxicol.* **17**, 16.

Hartwig, A., Arand, M., Epe, B., Guth, S., Jahnke, G., Lampen, A., Martus, H.-J., Monien, B., Rietjens, I. M. C. M., Schmitz-Spanke, S., *et al.* (2020). Mode of action-based risk assessment of genotoxic carcinogens. *Arch. Toxicol.* **94**, 1787–1877.

Health and Environmental Sciences Institute (HESI). (2022a). *Emerging Systems Toxicology for the Assessment of Risk (ESTAR). HESI - Health and Environmental Sciences Institute (blog)*. 2022. Available at: https://hesiglobal.org/emerging-systems-toxicology-for-assessment-of-risk-committee/.

Health and Environmental Sciences Institute (HESI). (2022b). *Problem Formulation – RISK2. Risk Assessment Formulation Tool (RAFT)*. 2022. Available at: https://risk21.org/problem-formulation/.

Hoshino, N., Tani, E., Wako, Y., and Takahashi, K. (2005). A two-generation reproductive toxicity study of benzophenone in rats. *J. Toxicol. Sci.* **30**, S5–S20.

International Cooperation on Cosmetic Regulation (ICCR). (2018). *Report on Integrated Strategies for Safety Assessment of Cosmetic Ingredients: Part 2*. Available at: https://www.iccrcosmetics.org/files/8315/4322/3079/ICCR_Integrated_Strategies_for_Safety_Assessment_of_Cosmetic_Ingredients_Part_2.pdf.

International Cooperation on Cosmetic Regulation (ICCR). (2021). *Report for the International Cooperation on Cosmetics Regulation Integrated Strategies for Safety Assessment of Cosmetic Ingredients: Workshop Report*. Available at: https://www.iccr-cosmetics.org//downloads/topics/2021%20iccr%20integrated%20strategies%20for%20safety%20assessment%20of%20cosmetic%20ingredients%20-%202019%20workshop%20report.pdf.

Joint FAO/WHO Expert Committee on Food Additives (JECFA). (2022). *Evaluations of JECFA Database*. 2022. Available at: https://apps.who.int/food-additives-contaminants-jecfa-database/.

Jeon, H.-K., Sarma, S. N., Kim, Y.-J., and Ryu, J.-C. (2008). Toxicokinetics and metabolisms of benzophenone-type UV filters in rats. *Toxicology* **248**, 89–95.

Joint FAO/WHO Meeting on Pesticide Residues (JMPR). (2022). *Plant Production and Protection Division: List of Pesticides Evaluated by JMPS*

and JMPR. 2022. Available at: https://www.fao.org/agriculture/crops/thematic-sitemap/theme/pests/lpe/en/.

Judson, R. S., Magpantay, F. M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., Xia, M., Huang, R., Rotroff, D. M., Filer, D. L., *et al.* (2015). Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicol. Sci.* **148**, 137–154.

Keller, D. A., Juberg, D. R., Catlin, N., Farland, W. H., Hess, F. G., Wolf, D. C., and Doerrer, N. G. (2012). Identification and characterization of adverse effects in 21st century toxicology. *Toxicol. Sci.* **126**, 291–297.

Kleinstreuer, N. C., Ceger, P. C., Allen, D. G., Strickland, J., Chang, X., Hamm, J. T., and Casey, W. M. (2016). A curated database of rodent uterotrophic bioactivity. *Environ. Health Perspect.* **124**, 556–562.

Krebs, A., van Vugt-Lussenburg, B. M. A., Waldmann, T., Albrecht, W., Boei, J., Ter Braak, B., Brajnik, M., Braunbeck, T., Brecklinghaus, T., Busquet, F., *et al.* (2020). The EU-ToxRisk method documentation, data processing and chemical testing pipeline for the regulatory use of new approach methods. *Arch. Toxicol.* **94**, 2435–2461.

Ladics, G. S., Price, O., Kelkar, S., Herkimer, S., and Anderson, S. (2021). A weight-of-the-evidence approach for evaluating, in lieu of animal studies, the potential of a novel polysaccharide polymer to produce lung overload. *Chem. Res. Toxicol.* **34**, 1430–1444.

Luijten, M., Wackers, P. F. K., Rorije, E., Pennings, J. L. A., and Heusinkveld, H. J. (2021). Relevance of in vitro transcriptomics for in vivo mode of action assessment. *Chem. Res. Toxicol.* **34**, 452–459.

Madden, J. C., Enoch, S. J., Paini, A., and Cronin, M. T. (2020). A review of in silico tools as alternatives to animal testing: Principles, resources and applications. *Altern. Lab. Anim.* **48**, 146–172.

Markossian, S., Grossman, A., Brimacombe, K., Arkin, M., Auld, D., Austin, C. P., Baell, J. *et al.* eds. (2004). *Assay Guidance Manual*. Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD). http://www.ncbi.nlm.nih.gov/books/NBK53196/.

Meek, M. E., Boobis, A., Cote, I., Dellarco, V., Fotakis, G., Munn, S., Seed, J., and Vickers, C. (2014). New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis: WHO/IPCS framework on mode of action/species concordance analysis. *J. Appl. Toxicol.* **34**, 1–18.

Meek, M. E. (Bette), and Lipscomb, J. C. (2015). Gaining acceptance for the use of in vitro toxicity assays and QIVIVE in regulatory risk assessment. *Toxicology* **332**, 112–123.

Miller, J. A., and Miller, E. C. (1975). Metabolic activation and reactivity of chemical carcinogens. *Mutat. Res.* **33**, 25–26.

Moreau, M., Mallick, P., Smeltz, M., Haider, S., Nicolas, C. I., Pendse, S. N., Leonard, J. A., Linakis, M. W., McMullen, P. D., Clewell, R. A., *et al.* (2022). Considerations for improving metabolism predictions for in vitro to in vivo extrapolation. *Front. Toxicol.* **4**, 894569.

Murk, A. J., Rijntjes, E., Blaauboer, B. J., Clewell, R., Crofton, K. M., Dingemans, M. M. L., Furlow, J. D., Kavlock, R., Köhrle, J., Opitz, R., *et al.* (2013). Mechanism-based testing strategy using in vitro approaches for identification of thyroid hormone disrupting chemicals. *Toxicol. Vitro* **27**, 1320–1346.

Nakagawa, Y., and Tayama, K. (2001). Estrogenic potency of benzophenone and its metabolites in juvenile female rats. *Arch. Toxicol.* **75**, 74–79.

Nakagawa, Y., and Tayama, K. (2002). Benzophenone-induced estrogenic potency in ovariectomized rats. *Arch. Toxicol.* **76**, 727–731.

National Toxicology Program. (2021). *Integrated Chemical Environment (ICE) v*. 3.5. 2021. Available at: https://ice.ntp.niehs.nih.gov/.

National Center for Advancing Translational Science (NCATS). (2022). *Tox21 Gateway*. 2022. Available at: https://tripod.nih.gov/tox/.

National Institute of Biomedical Innovation, Health and Nutrition (NIBIOHN). (2022). *こ の DB に つ い て | Open TG-GATEs*. 2022. Available at: https://toxico.nibiohn.go.jp/english/.

National Institute of Environmental Health Sciences (NIEHS). (2019). *2019 OHAT Handbook Update and Clarification Summary Document*. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookclarification-march2019_508.pdf.

Ning, J., Chen, L., and Rietjens, I. M. (2019). Role of toxicokinetics and alternative testing strategies in pyrrolizidine alkaloid toxicity and risk assessment; state-of-the-art and future perspectives. *Food Chem. Toxicol.* **131**, 110572.

Novak, A. L., Shaw, D. J., and Clutton, R. E. (2022). Animal welfare requirements in publishing guidelines. *Lab. Anim.* **56**, 561–575.

NSF International. (2021a). *NSF BENZOPHENONE - 2020 Benzophenone (CAS# 119-61-9) - Oral Risk Assessment Document*. 2021. Available at: https://www.techstreet.com/standards/nsf-benzophenone-2020?product_id=2222851.

NSF International. (2021b). *NSF/ANSI/CAN 61-2021 - Drinking Water System Components - Health Effects*. 2021. Available at: https://webstore.ansi.org/Standards/NSF/NSFANSI612021600.

NSF International. (2021c). *NSF/ANSI/CAN 600-2021 - Health Effects Evaluation and Criteria for Chemicals in Drinking Water*. 2021. Available at: https://webstore.ansi.org/Standards/NSF/NSFANSI6002021.

National Toxicology Program (NTP). (2018). *Integrative Screening Strategies for Neurotoxicity at the NTP*. National Institute of Environmental Health and Safety. doi: 10.22427/NTP-DATA-002-00062-0001-0000-1.

Organization for Economic Co-operation and Development (OECD). (1998). *OECD Principles on Good Laboratory Practice*. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. OECD. doi:10.1787/9789264078536-en.

Organization for Economic Co-operation and Development (OECD). (2005). *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment*. OECD Series on Testing and Assessment No. 34. OECD Publishing, Paris. Available at: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2005)14&doclanguage=en.

Organization for Economic Co-operation and Development (OECD). (2010). *Test No. 417: Toxicokinetics. OECD Guidelines for the Testing of Chemicals, Section 4*. OECD. doi:10.1787/9789264070882-en.

Organization for Economic Co-operation and Development (OECD). (2012). *OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters*. 2012. Available at: https://www.oecd.org/chemicalsafety/testing/oecdworkrelatedtoendocrinedisrupters.htm.

Organization for Economic Co-operation and Development (OECD). (2014). *OECD Adverse Outcome Pathway Knowledge Base*. 2014. Available at: https://www.oecd.org/chemicalsafety/launch-adverse-outcome-pathways-knowledge-base.htm.

Organization for Economic Co-operation and Development (OECD). (2017a). *Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA). OECD Series on Testing and Assessment*. OECD. doi:10.1787/44bb06c1-en.

Organization for Economic Co-operation and Development (OECD). (2017b). *Guidance Document for Describing Non-Guideline In Vitro Test*

Methods. *OECD Series on Testing and Assessment*. OECD. doi:10.1787/9789264274730-en.

Organization for Economic Co-operation and Development (OECD). (2018a). *Guidance Document on Good In Vitro Method Practices (GIVIMP)*. *OECD Series on Testing and Assessment*. OECD. doi:10.1787/9789264304796-en.

Organization for Economic Co-operation and Development (OECD). (2018b). *Guidance Document on Good In Vitro Method Practices (GIVIMP)*. *OECD Series on Testing and Assessment*. OECD. doi:10.1787/9789264304796-en.

Organization for Economic Co-operation and Development (OECD). (2019). *Guiding Principles on Good Practices for the Availability/Distribution of Protected Elements in OECD Test Guidelines. OECD Series on Testing and Assessment, No. 298*. Environment, Health and Safety, Environment Directorate, OECD. Available at: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV-JM-MONO(2019)14%20&doclanguage=en.

Organization for Economic Co-operation and Development (OECD). (2020). *Integrated Approaches to Testing and Assessment (IATA) - OECD*. 2020. Available at: https://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm#Project.

Organization for Economic Co-operation and Development (OECD). (2021a). *Guideline No. 497: Defined Approaches on Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/B92879a4-En.

Organization for Economic Co-operation and Development (OECD). (2021b). *The OECD QSAR Toolbox v 4.5*. 2021. Available at: https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm.

Organization for Economic Co-operation and Development (OECD). (2021c). *OECD Harmonised Template 201: Intermediate Effects - OECD*. 2021. Available at: https://www.oecd.org/ehs/templates/harmonised-templates-intermediate-effects.htm.

Organization for Economic Co-operation and Development (OECD). (2021d). *Guidance Document on the Characterisation, Validation and Reporting of Physiologically Based Kinetic (PBK) Models for Regulatory Purposes. OECD Series on Testing and Assessment, No. 331*. Environment, Health and Safety, Environment Directorate, OECD. Available at: https://www.oecd.org/chemicalsafety/risk-assessment/guidance-document-on-the-characterisation-validation-and-reporting-of-physiologically-based-kinetic-models-for-regulatory-purposes.pdf.

Organization for Economic Co-operation and Development (OECD). (2022a). *Case Study on the Use of an Integrated Approach for Testing and Assessment (IATA) for New Approach Methodology (NAM) for Refining Inhalation Risk Assessment from Point of Contact Toxicity of the Pesticide, Chlorothalonil. Series on Testing and Assessment No. 367. ENV/CBC/MONO(2022)31*. Available at: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/cbc/mono(2022)31&doclanguage=en.

Organization for Economic Co-operation and Development (OECD). (2022b). *OECD Test Guidelines for Chemicals - OECD*. 2022. Available at: https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm.

Ouedraogo, G., Alexander-White, C., Bury, D., Clewell, H. J., Cronin, M., Cull, T., Dent, M., Desprez, B., Detroyer, A., Ellison, C., *et al.* (2022). Read-across and new approach methodologies applied in a 10-step framework for cosmetics safety assessment – A case study with Parabens. *Regul. Toxicol. Pharmacol.* **132**, 105161.

Pain, G., Hickey, G., Mondou, M., Crump, D., Hecker, M., Basu, N., and Maguire, S. (2020). Drivers of and obstacles to the adoption of toxicogenomics for chemical risk assessment: insights from social science perspectives. *Environ. Health Perspect.* **128**, 105002.

Pamies, D. *et al.* (2021). Guidance document on good cell and tissue culture practice 2.0 (GCCP 2.0). *ALTEX* **39**, 30–70.

Pandey, A., Jatana, G. K., and Sonthalia, S. 2022. Cosmeceuticals. In *StatPearls*. StatPearls Publishing, Treasure Island (FL). http://www.ncbi.nlm.nih.gov/books/NBK544223/.

Parish, S. T., Aschner, M., Casey, W., Corvaro, M., Embry, M. R., Fitzpatrick, S., Kidd, D., Kleinstreuer, N. C., Lima, B. S., Settivari, R. S., *et al.* (2020). An evaluation framework for new approach methodologies (NAMs) for human health safety assessment. *Regul. Toxicol. Pharmacol.* **112**, 104592.

Paul Friedman, K., Gagne, M., Loo, L.-H., Karamertzanis, P., Netzeva, T., Sobanski, T., Franzosa, J. A., Richard, A. M., Lougee, R. R., Gissi, A., *et al.* (2020). Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. *Toxicol. Sci.* **173**, 202–225.

Pearce, R. G., Setzer, R. W., Strope, C. L., Sipes, N. S., and Wambaugh, J. F. (2017). HTTK: R package for high-throughput toxicokinetics. *J. Stat. Softw.* **79**, 1–26.

Pelkonen, O., and Vähäkangas, K. (1980). Metabolic activation and inactivation of chemical carcinogens. *J. Toxicol. Environ. Health.* **6**, 989–999.

Pelkonen, O. (2009). Comparison of metabolic stability and metabolite identification of 55 ECVAM/ICCVAM validation compounds between human and rat liver homogenates and microsomes - a preliminary analysis. *ALTEX* **26**, 214–222.

Pham, L. L., Sheffield, T. Y., Pradeep, P., Brown, J., Haggard, D. E., Wambaugh, J., Judson, R. S., and Paul Friedman, K. (2019). Estimating uncertainty in the context of new approach methodologies for potential use in chemical safety evaluation. *Curr. Opin. Toxicol.* **15**, 40–47.

Phillips, A. L., Hammel, S. C., Hoffman, K., Lorenzo, A. M., Chen, A., Webster, T. F., and Stapleton, H. M. (2018). Children's residential exposure to organophosphate ester flame retardants and plasticizers: Investigating exposure pathways in the TESIE study. *Environ. Int.* **116**, 176–185.

Pradeep, P., Patlewicz, G., Pearce, R., Wambaugh, J., Wetmore, B., and Judson, R. (2020). Using chemical structure information to develop predictive models for in vitro toxicokinetic parameters to inform high-throughput risk-assessment. *Comput. Toxicol.* **16**, 100136.

Pradeep, P., Paul Friedman, K., and Judson, R. (2020). Structure-based QSAR models to predict repeat dose toxicity points of departure. *Comput. Toxicol.* **16**, 100139.

PubChem. (2022a). *PubChem*. 2022. Available at: https://pubchem.ncbi.nlm.nih.gov/.

Pubchem. (2022b). *Tox21 - PubChem BioAssay - NCBI*. 2022. Available at: https://www.ncbi.nlm.nih.gov/pcassay?term=tox21.

Punt, A. (2017). Non-animal approaches for toxicokinetics in risk evaluations of food chemicals. *ALTEX* **34**, 501–514.

Punt, A. (2020). New approach methodologies (NAMs) for human-relevant biokinetics predictions. *ALTEX* **37**, 607–622.

Ramanarayanan, T., Szarka, A., Flack, S., Hinderliter, P., Corley, R., Charlton, A., Pyles, S., and Wolf, D. (2022). Application of a new approach method (NAM) for inhalation risk assessment. *Regul. Toxicol. Pharmacol.* **133**, 105216.

Rendic, S., and Peter Guengerich, F. (2012). Contributions of human enzymes in carcinogen metabolism. *Chem. Res. Toxicol.* **25**, 1316–1383.

Ring, C. L., Pearce, R. G., Woodrow Setzer, R., Wetmore, B. A., and Wambaugh, J. F. (2017). Identifying populations sensitive to

environmental chemicals by simulating toxicokinetic variability. *Environ. Int.* **106**, 105–118.

Rotroff, D. M., Dix, D. J., Houck, K. A., Knudsen, T. B., Martin, M. T., McLaurin, K. W., Reif, D. M., Crofton, K. M., Singh, A. V., Xia, M., *et al.* (2013). Using in vitro high throughput screening assays to identify potential endocrine-disrupting chemicals. *Environ. Health Perspect.* **121**, 7–14.

Rotroff, D. M., Wetmore, B. A., Dix, D. J., Ferguson, S. S., Clewell, H. J., Houck, K. A., Lecluyse, E. L., Andersen, M. E., Judson, R. S., Smith, C. M., *et al.* (2010). Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicol. Sci.* **117**, 348–358.

Rovida, C. (2021). NAM-supported read-across: From case studies to regulatory guidance in safety assessment. *ALTEX* **38**, 140–150.

Ruiz, P., and Fowler, B. A. (2015). Exposure assessment, forward and reverse dosimetry*. In *Handbook on the Toxicology of Metals*, 4th ed. (G. F. Nordberg, B. A. Fowler, and M. Nordberg, Eds.), Vol. 1, pp. 141–153. Elsevier, San Diego. doi:10.1016/B978-0-444-59453-2.00007-X.

Sauve-Ciencewicki, A., Davis, K. P., McDonald, J., Ramanarayanan, T., Raybould, A., Wolf, D. C., and Valenti, T. (2019). A simple problem formulation framework to create the right solution to the right problem. *Regul. Toxicol. Pharmacol.* **101**, 187–193.

Scientific Committee on Consumer Safety (SCCS). (2018). *The SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation: 10th Revision*. Publications Office, LU. https://data.europa.eu/doi/10.2875/77673.

Schenk, B., Weimer, M., Bremer, S., van der Burg, B., Cortvrindt, R., Freyberger, A., Lazzari, G., Pellizzer, C., Piersma, A., Schäfer, W. R., *et al.* (2010). The ReProTect feasibility study, a novel comprehensive in vitro approach to detect reproductive toxicants. *Reprod. Toxicol.* **30**, 200–218.

Schüttler, A., Jakobs, G., Fix, J. M., Krauss, M., Krüger, J., Leuthold, D., Altenburger, R., and Busch, W. (2021). Transcriptome-wide prediction and measurement of combined effects induced by chemical mixture exposure in zebrafish embryos. *Environ. Health Perspect.* **129**, 47006.

Schwartz, C. L., Christiansen, S., Vinggaard, A. M., Axelstad, M., Hass, U., and Svingen, T. (2019). Anogenital distance as a toxicological or clinical marker for fetal androgen action and risk for reproductive disorders. *Arch. Toxicol.* **93**, 253–272.

Sipes, N. S., Wambaugh, J. F., Pearce, R., Auerbach, S. S., Wetmore, B. A., Hsieh, J.-H., Shapiro, A. J., Svoboda, D., DeVito, M. J., and Ferguson, S. S. (2017). An intuitive approach for predicting potential human health risk with the Tox21 10k library. *Environ. Sci. Technol.* **51**, 10786–10796.

Solecki, R., Rauch, M., Gall, A., Buschmann, J., Kellner, R., Kucheryavenko, O., Schmitt, A., Delrue, N., Li, W., Hu, J., *et al.* (2019). Update of the DevTox data database for harmonized risk assessment and alternative methodologies in developmental toxicology: Report of the 9th Berlin Workshop on Developmental Toxicity. *Reprod. Toxicol.* **89**, 124–129.

Stucki, A. O., Barton-Maclaren, T. S., Bhuller, Y., Henriquez, J. E., Henry, T. R., Hirn, C., Miller-Holt, J., Nagy, E. G., Perron, M. M., Ratzlaff, D. E., *et al.* (2022). Use of new approach methodologies (NAMs) to meet regulatory requirements for the assessment of industrial chemicals and pesticides for effects on human health. *Front. Toxicol.* **4**, 964553.

Tennant, R. E., Guesné, S. J., Canipa, S., Cayley, A., Drewe, W. C., Honma, M., Masumura, K., Morita, T., Stalford, S. A., and Williams, R. V. (2019). Extrapolation of in vitro structural alerts for mutagenicity to the in vivo endpoint. *Mutagenesis* **34**, 111–121.

Testai, E., Bechaux, C., Buratti, F. M., Darney, K., Di Consiglio, E., Kasteel, E. E., Kramer, N. I., Lautz, L. S., Santori, N., Skaperda, Z.-V., *et al.* (2021). Modelling human variability in toxicokinetic and toxicodynamic processes using Bayesian meta-analysis, physiologically-based modelling and in vitro systems. *EFS3* **18**, doi: 10.2903/sp.efsa.2021.EN-6504.

Tirumala, M. G., Anchi, P., Raja, S., Rachamalla, M., and Godugu, C. (2021). Novel methods and approaches for safety evaluation of nanoparticle formulations: A focus towards in vitro models and adverse outcome pathways. *Front. Pharmacol.* **12**, 612659.

Tralau, T., Oelgeschläger, M., Kugler, J., Bloch, D., Braeuning, A., Burgdorf, T., Marx-Stoelting, P., Ritz, V., Schmeisser, S., Trubiroha, A., *et al.* (2021). A prospective whole-mixture approach to assess risk of the food and chemical exposome. *Nat. Food* **2**, 463–468.

Ulrey, A., Kolle, S., Landsiedel, R., and Hill, E. (2021). How a GIVIMP certification program can increase confidence in in vitro methods. *ALTEX* **38**, 316–318

United States Environmental Protection Agency (US EPA). (2011). Exposure Factors Handbook 2011 Edition (Final) [EPA Report]. (EPA/600/R-09/052F).

United States Environmental Protection Agency (US EPA). (2014a). *Framework for Human Health Risk Assessment to Inform Decision Making*. *EPA/100/R-14/001*. Available at: https://www.epa.gov/sites/default/files/2014-12/documents/hhra-framework-final-2014.pdf.

United States Environmental Protection Agency (US EPA). (2014b). *Framework for Ecological Risk Assessment. Other Policies and Guidance*. November 23, 2014. Available at: https://www.epa.gov/risk/framework-ecological-risk-assessment.

United States Environmental Protection Agency, Office of Research and Development (US EPA). (2015). *Toxicity Forecasting. Overviews and Factsheets*. August 21, 2015. Available at: https://www.epa.gov/chemical-research/toxicity-forecasting.

United States Environmental Protection Agency (US EPA). (2017). *Exploring ToxCast Data: Downloadable Data. Data and Tools*. November 1, 2017. Available at: https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data.

United States Environmental Protection Agency (US EPA). (2020). *Chemical Safety for Sustainability Strategic Research Action Plan 2019-2022. Other Policies and Guidance*. March 2, 2020. Available at: https://www.epa.gov/research/chemical-safety-sustainability-strategic-research-action-plan-2019-2022.

US EPA, OCSPP. (2020). *Chemicals Undergoing Risk Evaluation under TSCA. Overviews and Factsheets*. April 1, 2020. Available at: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/chemicals-undergoing-risk-evaluation-under-tsca.

United States Environmental Protection Agency (US EPA). (2021a). *New Approach Methods Work Plan (v2)*. December 2021. EPA/600/X-21/209. Available at: https://www.epa.gov/system/files/documents/2021-11/nams-work-plan_11_15_21_508-tagged.pdf

United States Environmental Protection Agency (US EPA). (2021b). *List of Alternative Test Methods and Strategies (or New Approach Methodologies [NAMs])*. https://www.epa.gov/sites/default/files/2021-02/documents/nams_list_second_update_2-4-21_final.pdf.

United States Environmental Protection Agency (US EPA). (2022a). *40 CFR Chapter I – Environmental Protection Agency*. 2022. Available at: https://www.ecfr.gov/current/title-40/chapter-I.

United States Environmental Protection Agency (US EPA). (2022b). *CompTox Chemicals Dashboard*. 2022. Available at: https://comptox.epa.gov/dashboard/.

United States Environmental Protection Agency National Center for Computational Toxicology (US EPA NCCT). (2022). *Toxicity Forecaster (ToxCast) In Vitro Assays. Assay Documentation for Non-guideline In Vitro Test Methods*. 2022. https://clowder.edap-cluster.com/files/6215520fe4b039b22c7a7836.

United States Food and Drug Administration (US FDA). (2018). *Physiologically Based Pharmacokinetic Analyses — Format and Content. Guidance for Industry*. Available at: https://www.fda.gov/media/101469/download.

United States Food and Drug Administration (US FDA). (2020a). *In Vitro Drug Interaction Studies – Cytochrome P450 Enzyme- and Transporter-Mediated Drug Interactions. Guidance for Industry*. Available at: https://www.fda.gov/media/134582/download.

United States Food and Drug Administration (US FDA). (2020b). *The Use of Physiologically Based Pharmacokinetic Analyses — Biopharmaceutics Applications for Oral Drug Product Development, Manufacturing Changes, and Controls Guidance for Industry. Draft Guidance* Available at: https://www.fda.gov/media/142500/download.

United States Food and Drug Administration (US FDA). (2021a). *Advancing New Alternative Methodologies at FDA*. January 2021 Report. Available at: https://www.fda.gov/media/144891/download.

United States Food and Drug Administration (US FDA). (2021b). *Focus Areas of Regulatory Science*. FDA, January. Available at: https://www.fda.gov/science-research/advancing-regulatory-science/focus-areas-regulatory-science.

United States Food and Drug Administration (US FDA). (2022). *Publications Co-Authored by FDA on Alternative Methods*. FDA, March. Available at: https://www.fda.gov/science-research/about-science-research-fda/publications-co-authored-fda-alternative-methods.

van der Zalm, A. J., Barroso, J., Browne, P., Casey, W., Gordon, J., Henry, T. R., Kleinstreuer, N. C., Lowit, A. B., Perron, M., Clippinger, A. J. (2022). A framework for establishing scientific confidence in new approach methodologies. *Arch. Toxicol.* **96**, 2865–2879.

Vries, R. (2021). Applying evidence-based methods to the development and use of adverse outcome pathways. *ALTEX* **38**, 336–347.

Wambaugh, J. F., Wetmore, B. A., Ring, C. L., Nicolas, C. I., Pearce, R. G., Honda, G. S., Dinallo, R., Angus, D., Gilbert, J., Sierra, T., *et al.* (2019). Assessing toxicokinetic uncertainty and variability in risk prioritization. *Toxicol. Sci.* **172**, 235–251.

Wang, X., Cirit, M., Wishnok, J. S., Griffith, L. G., and Tannenbaum, S. R. (2019). Analysis of an integrated human multiorgan microphysiological system for combined tolcapone metabolism and brain metabolomics. *Anal. Chem.* **91**, 8667–8675.

Wang, Z., Walker, G. W., Muir, D. C. G., and Nagatani-Yoshida, K. (2020). Toward a global understanding of chemical pollution: A first comprehensive analysis of national and regional chemical inventories. *Environ. Sci. Technol.* **54**, 2575–2584.

Watt, E. D., and Judson, R. S. (2018). Uncertainty quantification in ToxCast high throughput screening. *PLoS One* **13**, e0196963.

Westmoreland, C., Bender, H. J., Doe, J. E., Jacobs, M. N., Kass, G. E. N., Madia, F., Mahony, C., Manou, I., Maxwell, G., Prieto, P., *et al.* (2022). Use of new approach methodologies (NAMs) in regulatory decisions for chemical safety: Report from an EPAA deep dive workshop. *Regul. Toxicol. Pharmacol.* **135**, 105261.

Wetmore, B. A., Wambaugh, J. F., Allen, B., Ferguson, S. S., Sochaski, M. A., Setzer, R. W., Houck, K. A., Strope, C. L., Cantwell, K., Judson, R. S., *et al.* (2015). Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing. *Toxicol. Sci.* **148**, 121–136.

Wetmore, B. A., Wambaugh, J. F., Ferguson, S. S., Sochaski, M. A., Rotroff, D. M., Freeman, K., Clewell, H. J., Dix, D. J., Andersen, M. E., Houck, K. A., *et al.* (2012). Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicol. Sci.* **125**, 157–174.

World Health Organization (WHO). (2021). *The Public Health Impact of Chemicals: Knowns and Unknowns - Data Addendum for 2019*. 2021. Available at: https://www.who.int/publications-detail-redirect/WHO-HEP-ECH-EHD-21.01.

World Health Organization (WHO). (2022). *One Health*. 2022. Available at: https://www.who.int/news-room/questions-and-answers/item/one-health.

Wijeyesakere, S. J., Wilson, D., Auernhammer, T., Parks, A., Kovacs, D., and Marty, M. S. (2019). Hybrid machine-learning/SMARTS profiling model for mitochondrial inhibition. *Appl. Vitro Toxicol.* **5**, 196–204.

Wijeyesakere, S. J., Wilson, D. M., Settivari, R., Auernhammer, T. R., Parks, A. K., and Marty, M. S. (2018). Development of a profiler for facile chemical reactivity using the open-source Konstanz information miner. *Appl. Vitro Toxicol.* **4**, 202–213.

Wijeyesakere, S. J., Wilson, D. M., and Sue Marty, M. (2020). Prediction of cholinergic compounds by machine-learning. *Comput. Toxicol.* **13**, 100119.

Wilk-Zasadna, I., Bernasconi, C., Pelkonen, O., and Coecke, S. (2015). Biotransformation in vitro: An essential consideration in the quantitative in vitro-to-in vivo extrapolation (QIVIVE) of toxicity data. *Toxicology* **332**, 8–19.

Williams, A. J., Lambert, J. C., Thayer, K., and Dorne, J. L. C. M. (2021). Sourcing data on chemical properties and hazard data from the US-EPA CompTox chemicals dashboard: A practical guide for human risk assessment. *Environ. Int.* **154**, 106566.

Yamasaki, K., Takahashi, M., and Yasuda, M. (2005). Two-generation reproductive toxicity studies in rats with extra parameters for detecting endocrine disrupting activity: Introductory overview of results for nine chemicals. *J. Toxicol. Sci.* **30**, S1–S4.

Yamasaki, K., Takeyoshi, M., Yakabe, Y., Sawaki, M., Imatanaka, N., and Takatsuki, M. (2002). Comparison of reporter gene assay and immature rat uterotrophic assay of twenty-three chemicals. *Toxicology* **170**, 21–30.

Yoon, M., Blaauboer, B. J., and Clewell, H. J. (2015). Quantitative in vitro to in vivo extrapolation (QIVIVE): An essential element for in vitro-based risk assessment. *Toxicology* **332**, 1–3.

Zang, Q., Mansouri, K., Williams, A. J., Judson, R. S., Allen, D. G., Casey, W. M., and Kleinstreuer, N. C. (2017). In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. *J. Chem. Inf. Model.* **57**, 36–49.