

# Risk assessment in the 21st century: where are we heading?

Susy Brescia<sup>1,\*</sup>, Camilla Alexander-White<sup>2</sup>, Hequn Li<sup>3</sup>, Alex Cayley<sup>4</sup>

<sup>1</sup>Health & Safety Executive, Chemicals Regulation Division, Redgrave Court, Merton Road, Bootle, Merseyside L20 7HS, UK,

<sup>2</sup>MKTox & Co Ltd, c/o LB Group, 80 Compar Crescent, Ipswich, Suffolk IP2 0EH, UK,

<sup>3</sup>Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedfordshire MK44 1LQ, UK,

<sup>4</sup>Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds LS11, 5PS, UK

\*Corresponding author. Chemicals Regulation Division, Redgrave Court, Merton Road, Bootle, Merseyside L20 7HS, UK. Tel: +44 (0)7767 754408;

E-mail: susy.brescia@hse.gov.uk

Reliance on animal tests for chemical safety assessment is increasingly being challenged, not only because of ethical reasons, but also because they procrastinate regulatory decisions and because of concerns over the transferability of results to humans. New approach methodologies (NAMs) need to be fit for purpose and new thinking is required to reconsider chemical legislation, validation of NAMs and opportunities to move away from animal tests. This article summarizes the presentations from a symposium at the 2022 Annual Congress of the British Toxicology Society on the topic of the future of chemical risk assessment in the 21st century. The symposium included three case-studies where NAMs have been used in safety assessments. The first case illustrated how read-across augmented with some *in vitro* tests could be used reliably to perform the risk assessment of analogues lacking data. The second case showed how specific bioactivity assays could identify an NAM point of departure (PoD) and how this could be translated through physiologically based kinetic modelling in an *in vivo* PoD for the risk assessment. The third case showed how adverse-outcome pathway (AOP) information, including molecular-initiating event and key events with their underlying data, established for certain chemicals could be used to produce an *in silico* model that is able to associate chemical features of an unstudied substance with specific AOPs or AOP networks. The manuscript presents the discussions that took place regarding the limitations and benefits of these new approaches, and what are the barriers and the opportunities for their increased use in regulatory decision making.

**Key words:** NAMs; NGRA; safety assessment; AOPs; computational toxicology.

## Introduction

The last two decades have witnessed a biotechnology revolution, thanks to marked developments in computing and molecular biology, and the explosion of tools for assessing perturbations of test systems by chemicals and other stressors. This has led to rapid advancements in *in silico*, *in chemico* and *in vitro* methods for the assessment of the hazards and risks of chemicals, which fall under the term of new approach methodologies (NAMs).

NAMs are defined as any technology, methodology, approach or combination that can provide information on chemical hazard and risk assessment, and avoid the use of animals.<sup>1,2</sup> NAMs include *in silico* and modelling tools, artificial intelligence and machine learning systems, high-throughput data, bioactivity *in vitro* assays, toxicogenomics-based tests, *ex vivo* systems, read-across and grouping approaches, microphysiological systems (e.g. organs-on-the chip) and *in vitro* to *in vivo* extrapolation (IVIVE).<sup>3</sup>

Reliance on animal tests for chemical safety assessment is increasingly being challenged, not only because of ethical reasons that make these approaches unacceptable to large parts of society, but also because they procrastinate regulatory decisions and because of concerns over the transferability of results to humans.<sup>4,5</sup>

Therefore, NAMs have the potential to improve safety assessments by using more human-relevant tools through coverage of

key biological pathways or targets. Next-generation risk assessment (NGRA) provides a way to integrate NAM data from various sources into the decision-making process.

NAMs can mimic human biology and provide mechanistic information about how a chemical may cause toxicity in humans. The mechanistic information can be organized into adverse-outcome pathways (AOPs) and AOP networks.<sup>6,7</sup> The AOP framework connects through causal relationships an initial exposure event to a series of mechanistic and testable key events (KEs) and ultimately to a clinically relevant disease outcome. When AOPs are augmented with concentration- or dose-response data for the different KEs in the pathway, they can be converted into fully quantitative models of toxicity.<sup>8</sup> AOPs for numerous toxicity effects have already been developed.<sup>9</sup> However, as more and more AOPs and quantitative AOPs (qAOPs) for many more apical endpoints are established, it may be possible to predict adverse outcomes (AOs) and their safe doses in an increasingly expanded toxicological space, by measuring molecular or cellular events *in vitro* at lower levels of biological organization without whole animal testing.

Although *in vitro* testing within qualitative AOPs will permit prediction of complex hazard endpoints (hazard identification), qAOPs will additionally predict the associated safe doses for risk assessment. A powerful tool for converting a qualitative AOP in a

qAOP capable of establishing human-relevant guidance values is physiologically based kinetic (PBK) modelling, as it can facilitate quantitative IVIVE. PBK models combine information on exposure, physiology and chemical properties, describing the absorption, distribution, metabolism and excretion (ADME) processes of a chemical, to estimate time-dependent concentrations in plasma and tissues.<sup>10–12</sup> NOECs (No Observed Effect Concentrations) identified from *in vitro* tests measuring lower KEs in an qAOP can be converted into *in vivo* (human) NOAELs (No Observed Adverse Effect Levels) for the risk assessment using PBK models.

There are many advantages in moving to an NAM-based assessment paradigm: ethical (using less or no animals), regulatory (animal tests are banned for certain types of chemicals), economic (more cost-effective), scientific (more robust, reproducible and human-relevant results) and practical (faster assessments ensuring more rapid protection of human health from an increasing number of chemicals out there).<sup>13–15</sup>

However, despite the rapid advances in biotechnology and computational modelling, regulatory safety assessment of chemicals continues to rely heavily on *in vivo* testing in animals, particularly for higher tier hazard endpoints such as repeated dose toxicity and reproductive toxicity. In addition, to date, there has been a focus on study-by-study replacements of animal tests with NAMs.<sup>16</sup> For complex hazards, this strategy is unlikely to be successful, and attention should be given to the integration of different alternative approaches, including existing data, *in vitro* batteries, available human data and expert judgement in a weight-of-evidence (WoE) approach and using the Integrated-Approach-to-Testing-and-Assessment (IATA) methodology.<sup>17</sup>

Significant barriers slowing the pace of this paradigm shift in chemical safety assessment are still numerous. For certain chemicals, there are rigid animal-based data requirements set in law. In addition, for certain chemicals, the data generated have to be suitable for risk assessment but also hazard classification, the latter being based on apical toxicity effects/AOs (similar to human pathologies) detected in whole animal testing rather than *in vitro* bioactivity responses. The majority of NAMs are not validated at international level (e.g. by the Organisation for Economic Co-operation and Development—OECD). Validation ensures robustness, standardization, harmonization and mutual acceptance of data, but it is very time-consuming and the gold-standard animal test against which validation is performed has often never been validated.<sup>18</sup> The lack of validation hampers regulatory acceptance in many jurisdictions.

There is also a lack of knowledge in handling and interpreting new datasets from new technologies. Furthermore, there are significant limitations in our understanding of complex hazards and there is concern that new technologies do not cover the full toxicological space addressed by whole animal testing.

Still some NAMs are more promising than others and integrated approaches could be part of the solution. It is most likely that tiered approaches should be tried first and that NAMs should be used initially for prioritization and screening purposes. As knowledge grows, confidence in the use of NAMs could be bolstered by safe-harbour-approaches where assessments are generated using traditional methods and NAMs in parallel.<sup>19</sup>

A symposium on this topic was held in April 2022 as part of the British Toxicology Society Annual Congress (Newcastle, UK). The speakers and audience explored a number of case studies where NAMs featured heavily. The individual presentations and resulting discussion are summarized herein.

## A 10-step framework for NGRA applying new approach methods: principles and case-studies in cosmetics safety evaluation (Camilla Alexander-White, MKTox and Co Ltd)

The development of NAMs and NGRA have been a major subject of research and innovation in the cosmetics sector for two decades or more, driven largely by regulatory changes in Europe but also consumer opinion on the ethics regarding use of animals. The use of traditional approaches to generating toxicology data in animals was banned completely from 2013 March 11 for the purposes of supporting cosmetics safety assessment under Cosmetics Regulation (EC) No.1223/2009.

Many different types of *in vitro* assays and computational methods have been developed during these decades, and it has now reached the time when these NAMs can be applied to perform NGRA for cosmetic product ingredients.<sup>20</sup> The expectation is not to replace a traditional toxicology assay with an *in vitro* assay or even a suite of assays. The intent is to use new types of NAM evidence, in an integrated scientific way case-by-case, working to a set of principles, and with an understanding of human biological and toxicological pathways that could lead to adverse pathological outcomes in consumers. It will not be possible to cover the whole of human biology using NAMs, but that has always been the case with animal models too, where some known rodent toxicity pathways are not relevant to human biology. The aim is to perform a risk assessment, with uncertainty analysis, with sufficient confidence based on the best evidence that can be generated using human-relevant NAMs, in order to be able to take a decision on consumer safety.

In 2018, a task force met under the International Cooperation on Cosmetics Regulation (ICCR) where a set of principles (Fig. 1) were agreed that could underpin the use of NAMs in cosmetics risk assessment.<sup>13,14</sup> This is a human-relevant, exposure-led and hypothesis-driven approach that is designed to prevent harm.

To illustrate how these principles and NAM data generation could be used together with the concept of next-generation read-across (RAX), a tiered and iterative approach in a 10-step framework was developed that could in principle be used in a regulatory context<sup>21</sup> (Fig. 2).

As the 10-step framework begins with identifying and defining the exposure scenario, it is useful to make all reasonable efforts for refining consumer exposure for the NGRA. The risk assessor is often asked to consider a range of different cosmetic product types in an aggregated exposure assessment, when the same ingredient is used in multiple different product types. Exposure assessment is performed in a tiered manner, starting in Tier 1 with simple deterministic assumptions about how much product is used per day by a consumer for each product type (defined in SCCS notes of guidance, 2021, in Europe), and a defined % inclusion level of cosmetic ingredient. Probabilistic modelling tools have been developed to consider the habits and practices of product use by consumers, so as not to be excessively conservative in quantitative estimations of exposure, but Tier 2 use more data-driven and realistic models in calculating a suitable exposure metric<sup>21</sup> (Fig. 3). Tier 2+ models bring in the additional use of product occurrence data to consider the likelihood of a consumer using the ingredient in different product types each day. Outputs from all Tiers of probabilistic exposure modelling can be used as inputs into an NAM for exposure, with PBK modelling being used to estimate internal dose. When using NAM data to perform an *ab initio* NGRA (i.e. using *in vitro* and *in silico* data alone) and not the principle of read-across, cell-based assay concentrations as points of departure for bioactivity ( $POD_{NAM}$ ) are compared with

Principle 1: the overall goal is a human safety assessment  
 Principle 2: the assessment is exposure-led  
 Principle 3: the assessment is hypothesis-driven  
 Principle 4: the assessment is designed to prevent harm  
 Principle 5: the assessment follows an appropriate appraisal of all existing information  
 Principle 6: the assessment uses a tiered and iterative approach  
 Principle 7: the assessment uses robust and relevant methods and strategies  
 Principle 8: sources of uncertainty should be characterised and documented  
 Principle 9: the logic of the approach should be transparently and explicitly documented

Fig. 1. Nine principles of NGRA for cosmetic ingredients.<sup>13,14</sup>

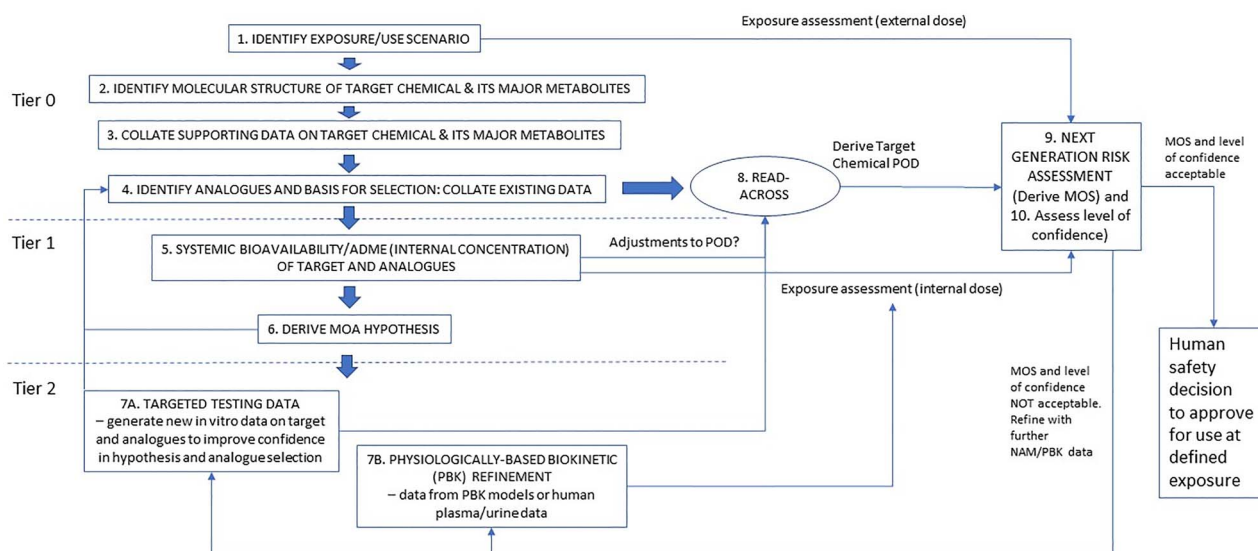


Fig. 2. A tiered 10-step framework to enable a human safety decision to be made using NAMs and next-generation RAX (using chemical and bioactivity data), which builds on the EU SEURAT-1 project workflow<sup>63</sup> to perform an NGRA without new animal data. [reproduced from 21].

- Tier 1 = highly conservative screening level – assumes presence in all products and daily use of all products.
- Tier 2 = moderately conservative – assumptions on product formulations are conservative
- Tier 2+ = remains conservative – assumptions on product formulations are more realistic but allow for fluctuations in ingredient use
- Tier 3 = measured data

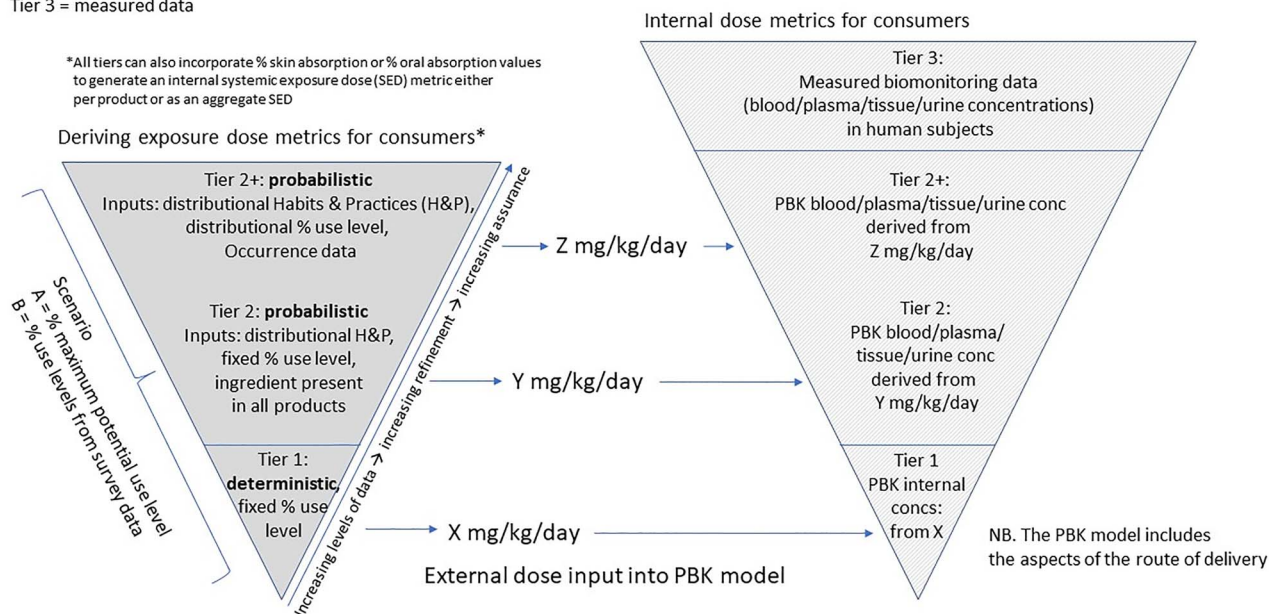


Fig. 3. The positioning of probabilistic exposure modelling and PBK when used in the context of refining exposure for a risk assessment.

internal systemic exposure concentrations (iSEC) in the human blood/plasma/target organs (see e.g. the *ab initio* case study for coumarin in Baltazar et al.,<sup>22</sup> and the approach to using NAMs in Middleton et al.<sup>23</sup>

To illustrate the implementation of the 10-step framework for NAM/RAX in NGRA in practice, two case studies have been published by the OECD, under the theme of IATA and in the peer review literature, for the safety assessment of parabens<sup>24,25</sup> and caffeine<sup>26,27</sup> as used in defined percentage inclusion levels in cosmetics products. The concept in these NAM/RAX case studies is that the “target substance” (either caffeine or propyl paraben, respectively) hypothetically has no data but that the target substance belongs to a family of similar substances for which there is already legacy toxicological data and existing points of departure for e similar analogues.

To develop an RAX hypothesis for NGRA (as per principle 3 above), it may be helpful to consider the scientific basis and definition of the RAX scenario on the basis of chemical similarity, metabolism and mode of action as follows<sup>28,29</sup>:

- (i) Chemical similarity of compounds that do not require (or do not undergo) metabolism to exert a potential adverse human health effect.
- (ii) Chemical similarity involving metabolism (resulting in exposure to the same/similar substance(s)).
- (iii) Chemicals with general low or no toxicity.
- (iv) Distinguishing chemicals (in a structurally similar category) with variable toxicities based on the MoA hypothesis.

In the case study for caffeine, a search for chemically similar analogues and existing knowledge on caffeine metabolism led to the decision to use the major caffeine metabolites as source chemicals for the target caffeine.<sup>26</sup> In the other case of parabens, the search for chemically similar analogues led to a focus on the homologous series of short-chain n-alkyl parabens, with the use of physico-chemical properties to help define the boundaries of similarity.<sup>24</sup> It is essential that the similarity hypothesis is pragmatic, can be supported by sufficient evidence to make it acceptable and fit for purpose, and, crucially, there are sufficient high-quality legacy toxicity data for the source molecules. To provide evidence of test substance and analogue similarity physico-chemical and ADME data, *in vitro cell-based assay* data and toxicogenomics data can be used.

In the case for an NGRA for propyl paraben, the selected POD was converted using PBK modelling to a maximum plasma concentration  $C_{max}$  representing a toxicological benchmark that was compared with a human equivalent  $C_{max}$  estimated following exposure to the ingredient in cosmetics use.<sup>24</sup> PBK is an essential tool going forward for NAM-based NGRA as is further illustrated in the next section.

### Quantitative *in vitro-in vivo* extrapolation)—bridging *in vitro* POD using PBK modelling. A case study (Hequn Li, Unilever)

Progress has been made over the last decades in the development of human-based *in vitro* NAMs to capture biological effects of compounds that may serve as non-animal methods for assuring human safety. However, translating *in vitro* effects into human health risks under real-world exposure conditions is challenging in NGRA as the kinetics of chemicals in the body is often ignored when interpreting *in vitro* effect data.<sup>13,30,31</sup> To bridge this data gap, *in silico* approaches, including PBK modelling that can facilitate quantitative *in vitro* to *in vivo* extrapolation, linking an

*in vitro*-derived biological (adverse) readout to a potential effect *in vivo*, are necessary. PBK models combine information on exposure, physiology and chemical properties, describing the absorption, distribution, metabolism and excretion (ADME) processes of a chemical, to estimate time-dependent concentration in plasma and tissues.<sup>10–12</sup>

PBK models have helped to quantitatively interpret *in vitro* effect concentrations of chemicals in the context of tissue dosimetry in many ways. For instance, they can be used to calculate internal exposure metrics such as  $C_{max}$  and AUC of the test chemical in plasma or tissues, which can help to identify tissues/organs with highest exposure or accumulation. These data can further guide the design and rationale of *in vitro* tests performed for risk assessment. Margins of safety (MoS) or bioactivity exposure ratio (BER) can be estimated by comparing key point of departure (PoD) concentrations derived from *in vitro* assays with human internal plasma or tissue concentrations for given external exposure situations predicted by PBK models. These MoS/BER values can be used for decision making, enabling the use of *in vitro* toxicity data for risk-based assessments beyond hazard-based assessments.<sup>22,23</sup> In addition, by simulating the plasma (or tissue) concentrations at different doses, one can infer the *in vivo* applied doses that are needed to reach the *in vitro* effect concentrations in the plasma (or tissue) (reverse dosimetry approach) and whether these effects are expected to be reached at defined exposure scenarios.<sup>32</sup>

Although PBK modelling is increasingly acknowledged for playing a crucial role in the toxicity testing paradigm shifts to non-animal approaches as a means of considering exposure and facilitating dosimetry, the development and gaining acceptance of the application of PBK models for NGRA decision making using NAMs remains a challenge. Two frequently asked questions are as follows: “How confident are we in the PBK model predictions used for risk assessment?” and “How can we identify and address the areas where we are least confident?”. The section below focuses on two key challenges we face today and potential solutions to overcome them.

### Challenge 1—PBK model parametrization partially or entirely based on data from *in vitro* or *in silico* methods

As we transition towards animal-free testing strategies for chemical safety evaluations, PBK model development has started to rely mostly or entirely on model parameters determined using NAMs. Over the past years, substantial advances have been made with the development and usage of *in vitro* or *in silico* methods for PBK model parametrization. It has been widely acknowledged and accepted that confidence in developed PBK models is highly dependent on the relevance and quality of the input parameter data. Recently, the OECD published a guidance document describing a scientific workflow for characterizing and validating PBK models developed using *in vitro* and *in silico* data.<sup>33</sup> The guidance document also provides an assessment framework for evaluating these models, with emphasis on the major uncertainties underlying model inputs and outputs.

Despite the efforts made in this area, challenges remain with respect to the availability and generation of sufficiently human relevant, robust and high-quality data for PBK model parametrization. Several frameworks for PBK modelling in NGRA using NAMs have been proposed,<sup>10,34,35</sup> and there is consensus amongst those frameworks that a tiered iterative approach needs to be applied for data generation and PBK model construction/refinement to ensure that the PBK models developed are fit for purpose and

provide sufficient confidence. Depending on the context, at a low tier, simplified compartmental models could be developed to develop an understanding of the expected order of magnitude of the internal concentrations for, e.g. chemical screening and prioritization purposes. At higher tiers, more refined and complex PBK models might be needed, e.g. when systemic concentrations e.g. a plasma or tissue  $C_{max}$  needs to be estimated to, combined with *in vitro* biological effect data, calculate MoS or BER values for decision making. In these instances, a pragmatic strategy would be to first build a PBK model with a generic model structure, then gradually increase complexity resulting in increasing extents of refinements. For an initial generic model, the minimal set of parameters needed are absorption rate, intrinsic hepatic clearance ( $Cl_{int}$ ), tissue: plasma partition coefficients, the fraction unbound in plasma (fup) and passive renal excretion (glomerular filtration rate times the fup). For data poor chemicals, the initial parametrization could rely solely on *in silico* predicted values, which can then be refined by progressively incorporating *in vitro* experimentally determined parameters. The requirements for data generation are guided at each step by sensitivity analysis together by hypotheses and rationale emerging from new data and new knowledge, increasing the confidence in model predictions.

Gaps and challenges as well remain with regards to ADME parameters generation, even though considerable efforts have been made to evaluate the quality of available *in vitro* or *in silico* models.<sup>36,37</sup> Many ADME parameters could be derived from multiple methods. For example, for determining hepatic  $Cl_{int}$ , one may consider using *in silico* QSAR models for prediction, or *in vitro* stability assays using suspension cultures of human hepatocytes, monolayer hepatocyte cultures, liver microsomes or liver S9 fraction. Each model has its own advantages and limitations. It is critical for the practitioner to have a good understanding of the mechanistic and biological relevance of the model of interest to the *in vivo* human situation. For instance, for low permeability chemicals, it is important to use a cell-based system to determine  $Cl_{int}$  so that the impact of the potentially slow rate of chemical movement across the cell membrane can be taken into consideration. Some higher tier or bespoke models of  $Cl_{int}$  may be considered such as a hepatocyte sandwich cultures for slowly metabolized chemicals, or transporter-certified hepatocytes depending on the chemicals of interest. In addition to multiple models, multiple protocols also exist to measure *in vitro* kinetic constants for metabolism. ADME parameter values obtained from different models or based on different protocols can differ by more than one order of magnitude, which could consequently lead to big differences in PBK prediction. This illustrates that there are still gaps to fill to achieve robust and reliable data generation that can be used in regulatory applications. To gain regulatory acceptance of *in vitro* kinetic data and PBK models based on *in vitro* input data, frameworks to allow harmonization of methodological aspects is required to avoid untargeted influences.<sup>38</sup> Another important aspect to consider when designing *in vitro* ADME assays are the chemical's physicochemical properties of the test material. Chemical instability, volatility and non-specific binding are factors that can impact the true exposure in *in vitro* test systems, therefore resulting in over- or underestimation of the effective concentrations both in ADME and toxicity assays. It is therefore very important to understand the potential influence of all these aspects in a case-by-case manner as all chemicals behave differently. Likewise, it is very important to report in a transparent way how the *in silico* and *in vitro* data were derived and measured to give confidence in the quality of the model.

## Challenge 2—PBK model evaluation in the absence of *in vivo* kinetic data

Traditionally, PBK model outputs are validated by comparison against human kinetic data. Confidence in the predictive capacity of a PBK model is based partly on the concordance between model predictions and *in vivo* kinetic data that has not been used to parametrize the model.<sup>33</sup> Currently, when *in vivo* data are missing, the confidence of the PBK models is often considered low. Two approaches could be considered/applied to fill this gap, which has become critical in increasing the confidence in using PBK modelling to help make safety decisions, which is especially important for regulatory acceptance.

One approach is to identify and quantify the uncertainty in PBK models in the absence of human kinetic data either through a bottom up (combining individual uncertainties of major sources into an overall modelling uncertainty) or top-down approach (holistic consideration of the overall model uncertainty). The uncertainty inherent to any PBK model can be divided into two categories: uncertainties in the model input parameters (i.e. variability in the parameter estimates due to intrinsic biological variation or measurement error) and uncertainty in the model algorithm/structure (biological basis). Apart from these intrinsic attributes, uncertainty can also arise from the level of experience of the modeller, different PBK platforms (software) used and data of different level of parametrization. Understanding the accuracy of PBK models and any tendencies to under- or overpredict systemic concentrations is important to establishing confidence. To gain this understanding, it would be helpful to explore the impact of different sources of uncertainty on PBK model confidence through case studies representing different modelling scenarios, by comparing outputs from the models to measured human kinetic data. The concept behind this approach is to first make PBK predictions for a list of chemicals/scenarios for which human kinetic data are available for qualification to calculate the degree of error for these scenarios. An assumption could then be made that for a new chemical without human kinetic data the degree of error for a certain modelling scenario is the same as previously established and therefore can be applied to account for uncertainty/prediction error. To achieve this, a database of available ADME data and human clinical kinetic data (e.g. a collection of plasma-concentration-time data, clearance, for individual study subjects in clinical trials with a detailed description of the study design/dose) covering a wide chemical space would be useful. PBK models could be built based on the available ADME data. Different sources of uncertainty would need to be covered including diverse physicochemical and ADME properties of the chemicals, different PBK platforms/software, different modellers, different level of parametrization (e.g. *in silico* only vs. *in vitro* derived values on generic parameters, bespoke PBK involving, e.g. active transporter) and different routes of administration (e.g. oral, i.v., dermal). Statistical approaches such as Bayesian modelling could then be applied to learn the statistical fold error representing various source of uncertainty from comparing PBK predictions of these scenarios with the observed human kinetic data. Depending on the data availability and chemical of interest, the overall fold errors obtained from the statistical model could be conducive to define an uncertainty factor for future chemicals in risk assessment decision making.

Another approach to confirm and build confidence in the reliability of PBK model predictions is through a PK analogue/read-across approach. The concept of this approach is that the validity of a PBK model developed for a target chemical (the chemical lacking *in vivo* kinetic data) could be evaluated using clinical

kinetic data from a source chemical (a chemical with existing kinetic data) based on the hypothesis that chemicals that have similar structure, physiochemical and/or ADME properties will have similar kinetics. Source chemicals could be identified based on either structural similarity and/or functional kinetics and ADME similarity. This PK analogue approach as an alternative method to evaluate PBK models has been demonstrated successfully in recent publications.<sup>39,40</sup> One of the challenges of this approach is the identification of appropriate read-across chemicals that shares structural or functional similarity. This becomes especially challenging when the target chemicals have some unique ADME properties, e.g. active transport, extrahepatic clearance, etc. The above-mentioned database on existing ADME data and human kinetic data becomes critical to an efficient application of this proposed PK analogue/read-across approach. Recent efforts have been made curating a range of ADME relevant data, existing PBK models and human kinetic data.<sup>41</sup> To have even better coverage of chemical types, more effort is required to gather more available data and more case studies/examples are needed to demonstrate the application of this approach.

To conclude, PBK modelling, which plays a key role in NGRA as a critical tool for IVIVE, is an area of active development. Further efforts are needed to establish a higher degree of confidence in the application of such models in a regulatory context.

## Integration of AOPs and computational models in chemical safety assessment (Alex Cayley, Lhasa limited)

### Introduction

Although NAMs can provide more human relevant results for safety assessment and minimize the use of animals, it is unlikely that a single alternative method will be able to replace the complexities of a whole animal model in isolation and, as a consequence, a method of interpreting and organizing the evidence being generated by these approaches is required to reach consistent, meaningful and transparent conclusions. AOPs represent a powerful framework to organize this knowledge and have been recognized by a number of international bodies as a good way to bring all of this evidence together to reach meaningful decisions.<sup>17</sup>

AOPs describe simplified biological pathways delineating a series of causally related events leading from the first interaction and perturbation of a biological system by a chemical, known as the molecular-initiating event (MIE), to the AO occurring at an individual or population level. The KEs between these two are captured at various different levels of biological complexity (molecular, cellular, tissue, organ) and are linked together by key event relationships.<sup>6</sup> One important aspect of the KEs is that they should be measurable and as a result each KE can be related to both assays and prediction models. These key features of AOPs allow for evidence to be contextualized on them in order that better safety decisions can be made. Indeed, this approach to using AOPs as a framework to interpret evidence has already been employed in the development of a defined approach for the assessment of skin sensitization<sup>42</sup> and is being used in the development of IATA for a number of other toxicity endpoints.<sup>43–47</sup>

### Elements required to make AOPs useful in chemical safety assessment

In order for AOPs to live up to their potential as a valuable tool in chemical safety assessment, a number of criteria must be met. Firstly, current knowledge around the biological processes leading

to toxicity need to be captured in the AOP format. Moreover, a comprehensive coverage of existing knowledge needs to be represented in this way and integrated into a coherent network, where the same concepts described in each separate AOP can be reused, built upon and joined together without repetition. With this knowledge in place, methods of testing it should then be associated with the appropriate places on the network, relevant data generated according to these test methods accessed, and an approach to reasoning between all of this evidence developed in order to reach meaningful conclusions. Finally, in order to make the approach reproducible, reusable, extensible and allow for scalability, the knowledge and links need to be digitized within a formal database structure that can be accessed through relevant software (Fig. 4).

### (A) Capture knowledge as AOPs

At first impression, the prospect of converting all known and emerging toxicological knowledge into AOPs may seem daunting, however, there are numerous different initiatives underway aimed at facilitating and focussing this process,<sup>9,48–51</sup> allowing for specific areas of toxicological space to be explored.

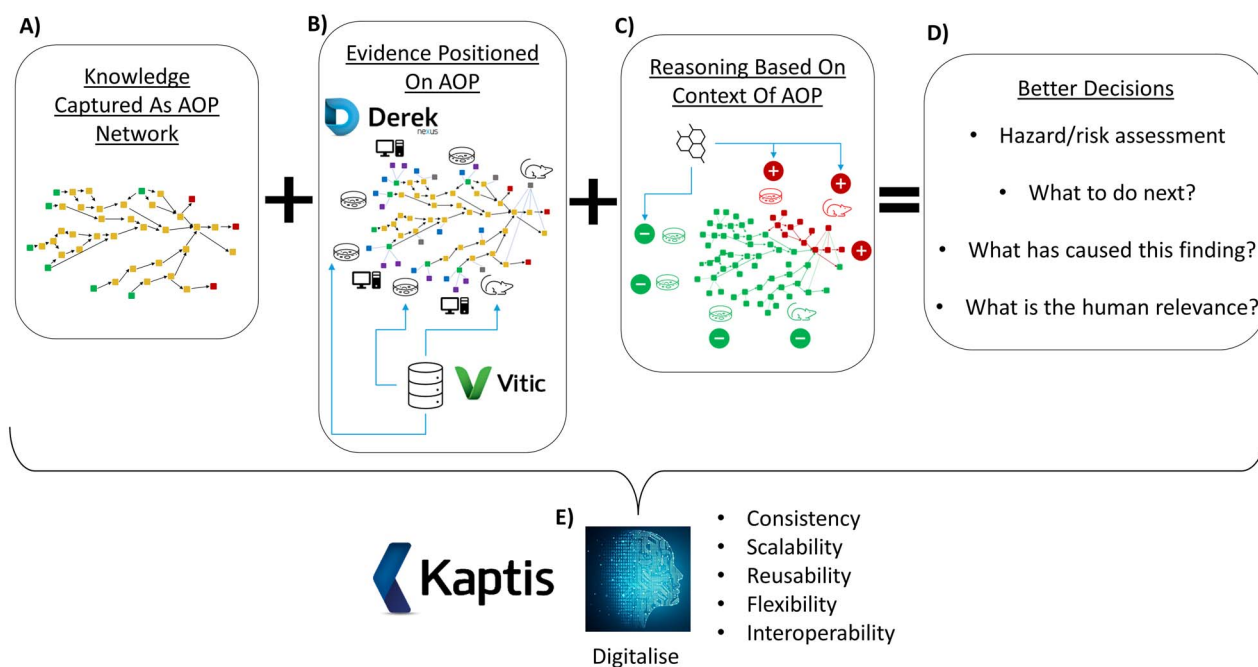
The work of Lhasa Limited in this area has, so far, focussed on two specific areas of toxicological space, carcinogenicity and developmental and reproductive toxicity (DART). These were chosen both as a result of the specific attention given to these endpoints in many safety assessment processes, meaning they could be well defined in terms of coverage, and also as a consequence of experience and knowledge in these domains held within the company. Projects were undertaken to harvest knowledge already captured in the predictive systems of Lhasa Limited,<sup>52</sup> as well as knowledge found in the public literature, in order to produce coherent AOPs and AOP networks relating to these two endpoints.<sup>53–57</sup> AOPs already captured within the public domain were also incorporated, ensuring that terms were standardized and integrated between pathways wherever possible to maintain the integrity of the network. To date, this has resulted in networks containing 38 MIEs relating to carcinogenicity<sup>54</sup> and 54 relating to DART.

### (B) Associate evidence

With a coherent network of AOPs covering a defined area of toxicological space having been developed, the next step is to associate evidence, in order that they can be used to contextualize this information. Without having evidence associated with them to test the various KEs, AOPs remain an academic construct and cannot fulfil their purpose of acting as a framework for decision making.

Although there are a number of public initiatives underway aimed at linking test methods and data with AOPs, the structure to the data being captured is currently rather limited or relies on public sources of knowledge and data which may not have been curated to ensure fidelity. Furthermore, these efforts generally do not consider evidence that can be obtained from *in silico* predictive systems and often the ability to associate the types of data formats generated by many new approach methods (NAMs) is limited.

Work at Lhasa limited has focussed on developing a formalized and structured approach to associating both assays and *in silico* predictions with their AOPs.<sup>58</sup> The extensive databases of toxicity data developed by Lhasa<sup>59</sup> have been used to link data within these assays and the first *in silico* predictions coming from the relevant alerts contained in the expert rule-based predictive system Derek Nexus<sup>60</sup> have been linked within the network. Over



**Fig. 4.** Organizing knowledge around an AOP framework. (A) Capturing knowledge as AOPs; (B) associating evidence and data with the appropriate places on these AOPs; (C) reasoning between evidence within the context of AOPs; (D) making better decisions based on this clarity and context; (E) digitizing the approach within a formalized database and exposing through the AOP software Kaptis.<sup>62</sup>

350 Derek alerts, over 100 assays and data associated with these assays for over 200 thousand compounds have been associated within the network up to this point. Work is currently underway to improve the connectivity of the network with emerging NAMs, including biomarker and transcriptomics approaches. The association of additional *in silico* models to the network is also being investigated.<sup>53</sup>

#### (C, D) Reason between evidence in the context of an AOP

Having captured knowledge as AOPs and associated evidence with them in the relevant places, the structure obtained can be used to compare evidence in more meaningful ways and make better decisions around chemical safety.

The questions which may be better answered by having knowledge structured in this way will include the obvious one of, “what safety decision should I take?” (whether this be a hazard- or risk-based decision) but might also include, “What should I do next?” or “Do I have enough evidence to make a decision?”, “what has caused this finding and what is its human relevance?” or even “what should I do next to improve my confidence in the result I am getting?”. Therefore, a method of reasoning between the evidence we have within the context of the AOP to answer these questions is required.

Research in this area has been undertaken at Lhasa Limited to establish methods by which this evidence might be combined in a meaningful way, ranging from a conservative combination of evidence on different KEs, to more sophisticated methods involving weighting of the different pieces of evidence and prioritizing their use based on the reliability, position on the AOP and association with the AO.<sup>58</sup>

#### (E) Digitize approach

The key to making AOPs and associated evidence scalable and therefore viable as a more general approach to facilitating safety assessment is in the digitization of this knowledge. Storing and

linking knowledge around AOPs in a well-structured database with enough granularity that they can be used to inform decisions allows the knowledge to be found, accessed and connected according to FAIR data principles.<sup>61</sup>

Producing a machine-readable knowledge base also means that the data and knowledge can be accessed quickly and in an automated fashion, allowing for a high throughput of queries of the knowledge and facilitating screening cases using the construct.

A database to capture this detailed knowledge surrounding AOPs, link the concepts available and the knowledge described previously has been developed by Lhasa Limited<sup>58</sup> The database is currently being used as the engine to drive the development of the software, Kaptis,<sup>62</sup> where it can be visualized, interrogated and used to reach safety decisions.

#### Developing software to enable the revolution

With detailed knowledge having been captured and contextualized around AOPs and this having been digitized within a database in the Kaptis software (according to Fig. 1), cases are currently being investigated, which would benefit from having evidence presented in this way in order to make better safety decisions.

One area which shows particular promise is in the field of carcinogenicity safety assessment. As discussed previously, many of the concepts relating to carcinogenicity have been captured within the Kaptis knowledge structure. A coherent network of known mechanisms leading to cancer have been developed, data associated and a simple reasoning approach is now being investigated.<sup>54</sup>

More specifically, an addendum to the current ICH S1 guidance,<sup>64</sup> relating to the assessment of pharmaceuticals for their carcinogenic potential, has recently been published.<sup>65</sup> This addendum describes an approach to carrying out an assessment based on a weight of evidence relating to existing evidence and data generated during the drug development process and using this to

reach a decision. This approach can save time, money and animals by avoiding the need to carry out a 2-year rat bioassay where evidence exists, which means this study would not add value.

The weight of evidence assessment advocated in the addendum seems particularly amenable to being represented in the context of AOPs, which will help with decision making in many ways.<sup>45</sup> First, a more consistent, transparent and interpretable presentation of the evidence available and the conclusions reached can be achieved. This will allow for more readily defensible submissions by the sponsor and means that the regulator and sponsor are more likely to reach a consensus in their conclusions based on the WoE available. In addition, as new evidence types become available in the form of NAMs, it will be easier to incorporate these and use them in the WoE.

Current work on the ICH S1 functionality within Kaptis will facilitate this process. A user can input a structure and associate a protein or gene target within the interface, the software then automatically looks up and associates relevant evidence with this query chemical by accessing the databases and models that have been connected. By linking this evidence with the KEs and AOPs within the network, the evidence can be presented within the context of the AOPs and preliminary automated conclusions reached at the various levels of the knowledge, including the KE, AO and overall AOP.

Although methods for combining evidence to reach automated conclusions, which can be used to answer questions are a key starting point in the output of Kaptis, it is also important that the decisions being reached and reasoning behind them are transparent and can be probed in order that the user can easily carry out expert review of any result that is being presented. The user can then interrogate the evidence underlying the conclusions and make arguments as to whether there is enough to decide the compound is likely to be a human carcinogen or not, such that a 2-year rat bioassay would not add value to the assessment. Alternatively, they may argue that the evidence is not conclusive (perhaps leading to the conclusion that more evidence should be gathered), in accordance with the addendum.

This functionality is currently being built within the software through a collaborative scientific consortium in order that the use case can be fulfilled. Work is also underway to extend the use of this software to other decisions around carcinogenicity and other endpoints.

## Concluding remarks

As illustrated by the case studies described above, chemical risk assessment is at a crossroad. Traditional safety assessments exclusively based on animal tests are becoming a part of the past, with the 21st century witnessing a gradual and steady increment in the more confident application of NAMs and alternatives to animal testing to chemical evaluations.

A full paradigm shift to animal-free assessments will not happen overnight as a full replacement of animal tests with NAMs is limited by specific legislative requirements, knowledge gaps, lack of validated alternatives, insufficient coverage of the toxicological space, inadequate infrastructure in contract research organizations for carrying out alternative methods, lack of education and limited confidence in NAMs for regulatory decision making.

Nevertheless, despite numerous barriers, opportunities to start paving a new way are already within reach:

- (i) Increased use of read-across/grouping by augmenting the validity of such an approach with the application of NAMs, e.g. omics-based technologies.
- (ii) Improved reliability of (Q)SAR ((quantitative) structure activity relationship) tools and predictions by the expansion of the chemical space through public access to privately owned data.
- (iii) Significant research investment in the creation of new AOPs, AOP networks and qAOPs for many more adverse effects, pathologies and apical endpoints, so that these can be predicted with their safe doses from earlier molecular and cellular events without whole animal testing.
- (iv) Development of *in silico* tools that can predict the AOPs triggered by certain chemicals without any testing.
- (v) Integration of different NAMs, different sources of information, including existing data in a WoE approach.
- (vi) Development of batteries of *in vitro* tests, covering different molecular and cellular pathways and phenotypes associated with complex toxicity endpoints.
- (vii) Acceleration of NAM validation at international level and creation of a tiered approach to validation where only methods having the greatest impact on regulatory decision making are fully validated.
- (viii) Investment in developing a wide range of bioactivity assays, including omics-based tests, capable of investigating an adequate number of stress responses associated with adverse effects and their concentration-response relationships.
- (ix) Further improvement of PBK models to translate *in vitro* NAM PODs to *in vivo* PODs.
- (x) Increased application of exposure-based waiving and exposure-driven testing.
- (xi) Improved application of artificial intelligence and machine learning approaches to increase predictivity and decision making.
- (xii) Further development of microphysiological systems (e.g. organs-on-the chip) to address complex endpoints.
- (xiii) Regulatory advocacy of using risk-based regulation for decision making that can make better use of NAMs in aiming for protection in certain exposure scenarios rather than prediction of hazard alone.

If progress is made in all of these areas, it will not be too long before some safety assessments could entirely be based on NAMs and *ab initio* alternative approaches. Confidence and acceptance for regulatory purposes will increase gradually as these approaches are applied in a step-wise manner first to the discovery phase of any new chemistry, then to screening and prioritization of ever more chemicals present out there, up to full regulatory decision making.

## Funding

Elements of this paper were funded by the Health and Safety Executive. In addition, the work on the 10-step framework for read across was funded by Cosmetics Europe.

## Conflicts of Interests statement

There are no conflicts of interest to declare. The authors' affiliation is as shown in the Title page. The views expressed here are the professional opinions of the authors and do not necessarily reflect those of the employing organizations. There are no contractual relations or proprietary considerations that restrict the authors publication or dissemination of the findings described in the manuscript.



## Author contributions

S.B.: conceptualization, co-ordination and supervision of manuscript, drafting, reviewing and editing of introduction and concluding remarks; C.A.-W.: investigations, drafting, review and editing of the first case study; H.L.: investigations, drafting, review and editing of the second case study; A.C.: investigations, drafting, review and editing of the third case study.

## References

- European Chemical Agency (ECHA). In: New approach methodologies in regulatory science. *Proceedings of a Scientific Workshop (19–20 April 2016)*. [https://echa.europa.eu/documents/10162/22816069/scientific\\_ws\\_proceedings\\_en.pdf](https://echa.europa.eu/documents/10162/22816069/scientific_ws_proceedings_en.pdf).
- EPA (Environmental Protection Agency). Strategic plan to promote the development and implementation of alternative test methods within the TSCA program. US Environmental Protection Agency. Office of Chemical Safety and Pollution Prevention, 22 June 2018:EPA-740-R1-8004.
- Krewski D, Andersen ME, Tyshenko MG, Krishnan K, Hartung T, Boekelheide K, Wambaugh JF, Jones D, Whelan M, Thomas R et al. Toxicity testing in the 21<sup>st</sup> century: progress in the past decade and future perspectives. *Arch Toxicol*. 2020;**94**(1):1–58.
- Hoffmann S, Edler L, Gardner I, Gribaldo L, Hartung T, Klein C, Liebsch M, Sauerland S, Schechtman L, Stammati A et al. Points of references in the validation process. *Altern Lab Anim*. 2008;**36**(3):343–352.
- Hoffmann S, Kleinstreuer N, Alépée N, Allen D, Api AM, Ashikaga T, Clouet E, Cluzel M, Desprez B, Gellatly N et al. Mom-animal methods to predict skin sensitisation (I): the Cosmetic Europe database. *Crit Rev Toxicol*. 2020;**48**(5):344–358.
- Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK et al. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem*. 2010;**29**(3):730–741.
- Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, Rowlands JC, Whelan MP, Hays SM, Andersen ME et al. Incorporating new technologies into toxicity testing and risk assessment: moving from 21<sup>st</sup> century vision to a data-driven framework. *Toxicol Sci*. 2013;**136**(1):4–18.
- Spînu N, Cronin MTD, Madden JC, Worth AP. A matter of trust: learning lessons about causality will make qAOPs credible. *Comput Toxicol*. 2022;**21**:100205.
- AOPwiki, aopwiki. <https://aopwiki.org/> (2021 July 16, date last accessed).
- Moxon TE, Li H, Lee MY, Piechota P, Nicol B, Pickles J, Pendlington R, Sorrell I, Baltazar MT. Application of physiologically based kinetic (PBK) modelling in the next generation risk assessment of dermally applied consumer products. *Toxicol In Vitro*. 2020;**63**:104746.
- Li H, Reynolds J, Sorrell I, Sheffield D, Pendlington R, Cubberley R, Nicol B. PBK modelling of topical application and characterisation of the uncertainty of C<sub>max</sub> estimate: a case study approach. *Toxicol Appl Pharmacol*. 2022;**442**:115992.
- Wetmore BA, Wambaugh JF, Allen B, Ferguson SS, Sochaski MA, Setzer RW, Houck KA, Strobe CL, Cantwell K, Judson RS et al. Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing. *Toxicol Sci*. 2015;**148**(1):121–136.
- Dent MP, Amaral RT, Da Silva PA, Ansell J, Boisleve F, Hatao M, Hirose A, Kasai Y, Kern P, Kreiling R et al. Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. *Comput Toxicol*. 2021a;**7**:20–26.
- Dent MP, Vaillancourt E, Thomas RS, Carmichael PL, Ouedraogo G, Kojima H, Barroso J, Ansell J, Barton-Maclaren TS, Bennekou SH et al. Paving the way for application of next generation risk assessment to safety decision-making for cosmetic ingredients. *Regul Toxicol Pharmacol*. 2021b;**125**:105026 Epub 2021 Aug 10. PMID: 34389358; PMCID: PMC8547713.
- Thomas RS, Bahadori T, Buckley TJ, Cowden J, Deisenroth C, Dionisio KL, Frithsen JB, Grulke CM, Gwinn MR, Harrill JA et al. The next generation blueprint of computational toxicology at the US Environmental Protection Agency. *Toxicol Sci*. 2019;**169**(2):317–332.
- Ball N, Bars R, Botham PA, Cuciureanu A, Cronin MTD, Doe JE, Dudzina T, Gant TW, Leist M, van Ravenzwaay B. A framework for chemical safety assessment incorporating new approach methodologies within REACH. *Arch Toxicol*. 2022;**96**(3):743–766.
- OECD (Organisation for Economic Co-operation and Development). Guidance document for the use of adverse outcome pathways in developing Integrated Approaches to Testing and Assessment (IATA): OECD series on testing and assessment No 260; 19 December 2016. (ENV/JM/MONO(2016)67).
- OECD (Organisation for Economic Co-operation and Development). 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; 18 August 2005. (ENV/JM/MONO(2005)14).
- Sewell F, Doe J, Gellatly N, Ragan I, Burden N. Steps towards the international regulatory acceptance of non-animal methodology in safety assessment. *Regul Toxicol Pharmacol*. 2017;**89**:50–56.
- Carmichael PL, Baltazar MT, Cable S et al. Ready for regulatory use: NAMs and NGRA for chemical safety assurance. *ALTEX*. 2022;**39**(3):359–366. Epub 2022 Jul 4 PMID: 35796331.
- Alexander-White C, Bury D, Cronin M, Dent M, Hack E, Hewitt NJ, Kenna G, Naciff J, Ouedraogo G, Schepky A et al. A 10-step framework for use of read-across (RAX) in next generation risk assessment (NGRA) for cosmetics safety assessment. *Regul Toxicol Pharmacol*. 2022;**129**:105094.
- Baltazar MT, Cable S, Carmichael PL, Cubberley R, Cull T, Dela-grange M, Dent MP, Hatherell S, Houghton J, Kucic P et al. A next-generation risk assessment case study for Coumarin in cosmetic products. *Toxicol Sci*. 2020;**176**(1):236–252.
- Middleton AM, Reynolds J, Cable S, Baltazar MT, Li H, Bevan S, Carmichael PL, Dent MP, Hatherell S, Houghton J et al. Are non-animal systemic safety assessments protective? A toolbox and workflow. *Toxicol Sci*. 2022;**189**(1):124–147.
- Ouedraogo G, Alexander-White C, Bury D, Clewell HJ III, Cronin M, Cull T, Dent M, Desprez B, Detroyer A, Ellison C et al. Read-across and new approach methodologies applied in a 10-step framework for cosmetics safety assessment—a case study with parabens. *Regul Toxicol Pharmacol*. 2022;**132**:105161 Epub 2022 May 1. PMID: 35508214.
- OECD (Organisation for Economic Co-operation and Development) IATA case studies project. Case Study on Use of an Integrated Approach to Testing and Assessment (IATA) and New Approach Methods to Inform a Theoretical Read-Across for Dermal Exposure to Propylparaben from Cosmetics. Series on Testing & Assessment No. 320; 2020. (ENV/JM/MONO(2020a)16).
- Bury D, Alexander-White C, Clewell HJ 3rd, Cronin M, Desprez B, Detroyer A, Efremenko A, Firman J, Hack E, Hewitt NJ et al. New framework for a non-animal approach adequately assures the safety of cosmetic ingredients—a case study on caffeine. *Regul Toxicol Pharmacol*. 2021b;**123**:104931. Epub 2021 Apr 24. 33905778.

27. OECD (Organisation for Economic Co-operation and Development). IATA case studies project. Case Study on Use of an Integrated Approach to Testing and Assessment (IATA) for Systemic Toxicity Arising from Cosmetic Exposure to Caffeine. Series on Testing & Assessment No. 321; 2020. (ENV/JM/MONO(2020b)17).
28. Berggren E, Amcoff P, Benigni R, Blackburn K, Carney E, Cronin M, Deluyker H, Gautier F, Judson RS, Kass GEN et al. Chemical safety assessment using read-across: assessing the use of novel testing methods to strengthen the evidence base for decision making. *Environ Health Perspect*. 2015;**123**(12):1232–1240.
29. European Chemical Agency (ECHA). Read-across assessment framework (RAAF); 2017. ECHA-17-R-01-EN, 978-92- 9495-758-0.
30. Westmoreland C, Carmichael P, Dent M, Fentem J, MacKay C, Maxwell G, Pease C, Reynolds F. Assuring safety without animal testing: Unilever's ongoing research programme to deliver novel ways to assure consumer safety. *Altern Lab Anim*. 2010;**27**(3):207–211.
31. National Research Council. *Toxicity Testing in the 21st Century: A Vision and A Strategy*. Washington, DC: The National Academies Press; 2007.
32. Li H, Yuan H, Middleton A, Li J, Nicol B, Carmichael P, Guo J, Peng S, Zhang Q. Next generation risk assessment (NGRA): bridging in vitro points-of-departure to human safety assessment using physiologically-based kinetic (PBK) modelling—a case study of doxorubicin with dose metrics considerations. *Toxicol In Vitro*. 2021a;**74**:105171.
33. OECD (Organisation for Economic Co-operation and Development). *Guidance Document on the Characterisation, Validation and Reporting of Physiologically Based Kinetic (PBK) Models for Regulatory Purposes*. Paris: OECD Publishing; 2021.
34. Li H et al. Generic PBK modelling of topical application and characterisation of the distribution of Cmax estimate errors: a case study approach. Submitted; 2021b.
35. Paini A, Leonard JA, Joossens E, Bessems JGM, Desalegn A, Dorne JL, Gosling JP, Heringa MB, Klaric M, Kliment T et al. Next generation physiologically based kinetic (NG-PBK) models in support of regulatory decision making. *Comput Toxicol*. 2019;**9**:61–72.
36. Bessems JG, Loizou G, Krishnan K, Clewell HJ III, Bernasconi C, Bois F, Coecke S, Collnot EM, Diembeck W, Farcal LR et al. PBTK modelling platforms and parameter estimation tools to enable animal-free risk assessment: recommendations from a joint EPAA–EURL ECVAM ADME workshop. *Regul Toxicol Pharmacol*. 2014;**68**(1):119–139.
37. Patel M, Chilton ML, Sartini A, Gibson L, Barber C, Covey-Crump L, Przybylak KR, Cronin MTD, Madden JC. Assessment and reproducibility of quantitative structure-activity relationship models by the nonexpert. *J Chem Inf Model*. 2018;**58**(3):673–682.
38. Lousse J, Alewijn M, Peijnenburg AACM, Cnubben NHP, Heringa MB, Coecke S, Punt A. Towards harmonization of test methods for in vitro hepatic clearance studies. *Toxicol In Vitro*. 2020;**63**:104722.
39. Ellison CA, Wu S. Application of structural and functional pharmacokinetic analogs for physiologically based pharmacokinetic model development and evaluation. *Regul Toxicol Pharmacol*. 2020;**114**:104667.
40. Paini A, Worth A, Kulkarni S, Ebbrell D, Madden J. Assessment of the predictive capacity of a physiologically based kinetic model using a read-across approach. *Comput Toxicol*. 2021;**18**:100159.
41. Thompson CV, Firman JW, Goldsmith MR, Grulke CM, Tan YM, Paini A, Penson PE, Sayre RR, Webb S, Madden JC. A systematic review of published physiologically-based kinetic models and an assessment of their chemical space coverage. *Altern Lab Anim*. 2021;**49**(5):197–208.
42. OECD (Organisation for Economic Co-operation and Development), OECD Guideline No. 497: defined approaches on skin sensitisation. 2021, <https://www.oecd.org/chemicalsafety/guideline-no-497-defined-approaches-on-skin-sensitisation-b92879a4-en.htm> (last accessed 2021 September 27).
43. Jacobs MN, Colacci A, Louekari K, Luijten M, Hakkert BC, Paparella M, Vasseur P. International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances. *ALTEX*. 2016;**33**(4):359–392 PMID: 27120445.
44. Jacobs MN, Colacci A, Corvi R, Vaccari M, Aguila MC, Corvaro M, Delrue N, Desaulniers D, Ertych N, Jacobs A et al. Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens. *Arch Toxicol*. 2020;**94**(8):2899–2923.
45. Stalford SA, Cayley AN, de Oliveira AAF. Employing an adverse outcome pathway framework for weight-of-evidence assessment with application to the ICH S1B guidance addendum. *Reg Toxicol Pharmacol*. 2021;**127**:105071. <https://www.sciencedirect.com/science/article/pii/S0273230021002129>.
46. European Chemical Agency (ECHA) and European Food Safety Authority (EFSA) with the technical support of the Joint Research Centre (JRC). Guidance for the identification of endocrine disruptors in the context of regulations (EU) No 528/2012 and (EC) No 1107/2009. *EFSA J*. 2018;**16**(6):e05311 PMID: 32625944; PMCID: PMC7009395.
47. Nymark P, Karlsson HL, Halappanavar S, Vogel U. Adverse outcome pathway development for assessment of lung carcinogenicity by nanoparticles. *Front Toxicol*. 2021;**3**:653386.
48. Arnesdotter E, Spinu N, Firman J, Ebbrell D, Cronin MTD, Vanhaecke T, Vinken M. Derivation, characterisation and analysis of an adverse outcome pathway network for human hepatotoxicity. *Toxicology*. 2021;**459**:152856. Epub 2021 Jul 10 PMID: 34252478.
49. Sasaki JC, Allemang A, Bryce SM, Custer L, Dearfield KL, Dietz Y, Elhajouji A, Escobar PA, Fornace AJ Jr, Froetschl R et al. Application of the adverse outcome pathway framework to genotoxic modes of action. *Environ Mol Mutagen*. 2020;**61**(1):114–134 PMID: 31603995.
50. Lynch AM, Eastmond D, Elhajouji A, Froetschl R, Kirsch-Volders M, Marchetti F, Masumura K, Pacchierotti F, Schuler M, Tweats D. Targets and mechanisms of chemically induced aneuploidy. Part 1 of the report of the 2017 IWGT workgroup on assessing the risk of aneuploids for carcinogenesis and hereditary diseases. *Mutat Res*. 2019;**847**:403025 PMID: 31699346.
51. Noyes PD, Friedman KP, Browne P, Haselman JT, Gilbert ME, Hornung MW, Barone S Jr, Crofton KM, Laws SC, Stoker TE et al. Evaluating chemicals for thyroid disruption: opportunities and challenges with in vitro testing and adverse outcome pathway approaches. *Environ Health Perspect*. 2019;**127**(9):95001. Epub 2019 Sep 5 PMID: 31487205; PMCID: PMC6791490.
52. Lhasa Limited. <https://www.lhasalimited.org/>.
53. Myden A, Hill E, Fowkes A. Using adverse outcome pathways to contextualise (Q)SAR predictions for reproductive toxicity—a case study with aromatase inhibition. *Reprod Toxicol*. 2022;**108**:43–55. Epub 2022 Jan 25 PMID: 35091028.
54. Cayley AN, Foster RS, Hill E, Kane S, Kocks G, Myden A, Newman D, Stalford SA, Vessey JD, Zarei R et al. Development of a network of carcinogenicity adverse outcome pathways and its employment as an evidence framework for safety assessment. *ALTEX* Epub ahead of print. 2022: PMID: 35575642.
55. Heusinkveld H, Braakhuis H, Gommans R, Botham P, Corvaro M, van der Laan JW, Lewis D, Madia F, Manou I, Schorsch F

- et al. Towards a mechanism-based approach for the prediction of nongenotoxic carcinogenic potential of agrochemicals. *Crit Rev Toxicol*. 2020;**50**(9):725–739 PMID: 33236972.
56. Johansson HKL, Damdimopoulou P, van Duursen MBM, Boberg J, Franssen D, de Cock M, Jäger K, Wagner M, Velthut-Meikas A, Xie Y et al. Putative adverse outcome pathways for female reproductive disorders to improve testing and regulation of chemicals. *Arch Toxicol*. 2020;**94**(10):3359–3379.
57. Janowska-Sejda EI, Adeleye Y, Currie RA. Exploration of the DARTable genome- a resource enabling data-driven NAMs for developmental and reproductive toxicity prediction. *Front Toxicol*. 2022;**3**:806311 PMID: 35295108 PMCID: PMC8915813.
58. Ball T, Barber CG, Cayley A, Chilton ML, Foster R, Fowkes A, Heghes C, Hill E, Hill N, Kane S et al. Beyond adverse outcome pathways: making toxicity predictions from event networks, SAR models, data and knowledge. *Toxicol Res*. 2021;**10**(1):102–122 PMID: 33613978.
59. Vitic. <https://www.lhasalimited.org/products/vitic.htm/>.
60. Derek Nexus. <https://www.lhasalimited.org/products/derek-nexus.htm/>.
61. Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;**3**(1):160018.
62. Kaptis. <https://www.lhasalimited.org/products/kaptis.htm>.
63. Berggren E, White A, Ouedraogo G, Paini A, Richarz AN, Bois FY, Exner T, Leite S, Grunsven, Worth A et al. Ab initio chemical safety assessment: a workflow based on exposure considerations and non-animal methods. *Comput Toxicol*. 2017;**4**:31–44.
64. ICH S1, International Council for Harmonisation (ICH) of technical requirements for Pharmaceuticals for Human Use, ICH Harmonised Guideline S1B, 1997, <https://database.ich.org/sites/default/files/S1B%20Guideline.pdf>.
65. ICH S1(R1), International Council for Harmonisation (ICH) of Technical Requirements for Pharmaceuticals for Human Use, ICH Harmonised Guideline S1B(R1), 2022, [https://database.ich.org/sites/default/files/S1B-R1\\_FinalGuideline\\_2022\\_0719.pdf](https://database.ich.org/sites/default/files/S1B-R1_FinalGuideline_2022_0719.pdf).