



**REPORT FOR THE  
INTERNATIONAL COOPERATION ON COSMETICS REGULATION**

***INTEGRATED STRATEGIES FOR SAFETY ASSESSMENT OF COSMETIC INGREDIENTS: WORKSHOP  
REPORT***

**Authors:**

Dent, Matt (9)#, Kojima, Hajime (4)#, Ansell, Jay (12), Barroso, João (3), Barton-Maclaren, Tara S. (2)+, Bennekou, Susanne H. (14)+, Boekelheide, Kim (18)+, Carmichael, Paul L. (9)+, Ezendam, Janine (15)+, Field, John (2), Fitzpatrick, Suzanne (5)+, Hatao, Masato (11), Kreiling, Reinhard (10), Lorencini, Marcio (6), Mahony, Catherine (9)+, Mazaro-Costa, Renata (13)+, Montemayor, Beta (7), Oliveira, Julcemara (1), Ouédraogo, Gladys (9), Rogiers, Vera (16)+, Smegal, Deborah (5)+, Taalman, Rob (9), Thomas, Russell S. (20)+, Tokura, Yoshiki (17)+, Vaillancourt, Eric (2), Verma, Rajeshwar (5), Willett, Catherine (19)+, Yang, Chien (8)

1. Brazilian Health Regulatory Agency (ANVISA), Brazil
2. Health Canada (HC), Canada
3. European Commission (EC), Europe
4. National Institute of Health Sciences (NIHS), Japan
5. US Food and Drug Administration (US FDA), United States
6. Brazilian Association of the Cosmetic, Toiletry and Fragrance Industry (ABIPHEC), Brazil
7. Cosmetics Alliance Canada (CAC), Canada
8. Taiwan Cosmetic Industry Association (TWCIA), Chinese Taipei
9. Cosmetics Europe (CE), Europe
10. European Federation for Cosmetic Ingredients (EFfCI), Europe
11. Japan Cosmetic Industry Association (JCIA), Japan
12. US Personal Care Products Council (PCPC), United States
13. Federal University of Goiás (UFG), Brazil
14. National Food Institute, Technical University of Denmark (DTU), Denmark, Europe
15. National Institute for Public Health and the Environment (RIVM), the Netherlands, Europe
16. Vrije Universiteit Brussel (VUB), Belgium, Europe
17. Chutoen General Medical Center, Japan
18. Brown University (BU), United States
19. Humane Society International (HSI), United States
20. U.S. Environmental Protection Agency (EPA), United States

# Co-chairs

+ non-JWG member, invited subject matter expert

## TABLE OF CONTENTS

|   |    |
|---|----|
| 1. SELECTED ABBREVIATIONS AND DEFINITIONS .....   | 3  |
| 2. BACKGROUND.....  | 4  |
| 3. PURPOSE .....  | 5  |
| 4. INTRODUCTION.....  | 6  |
| 5. THE LEVEL OF CONSUMER PROTECTION OFFERED BY NGRA .....   | 7  |
| 6. COMMON FEATURES IN NGRA FOR COSMETIC-RELEVANT CASE STUDIES .....   | 8  |
| 6.1. Application of ICCR Principles to evaluation of local toxicity .....   | 9  |
| 6.2. Systemic toxicity case studies .....   | 10 |
| 6.2.1. Overview of coumarin case study.....   | 10 |
| 6.2.2. Overview of phenoxyethanol case study.....   | 12 |
| 6.3. Application of ICCR Principles for the evaluation of systemic toxicity .....   | 14 |
| 6.3.1. The overall goal is a human safety risk assessment.....  | 14 |
| 6.3.2. The assessment is exposure led.....  | 15 |
| 6.3.3. The assessment is hypothesis drive .....   | 15 |
| 6.3.4. The assessment is designed to prevent harm.....  | 16 |
| 6.3.5. Conduct an appropriate appraisal of existing information.....  | 17 |
| 6.3.6. Use a tiered and iterative approach .....  | 17 |
| 6.3.7. Use robust and relevant methods and strategies.....  | 18 |
| 6.3.8. Characterize and document sources of uncertainty.....  | 18 |
| 6.3.9. Document the logic of the approach transparently .....   | 19 |
| 6.4. Common tools and approaches in the evaluation of systemic toxicity .....   | 19 |
| 6.4.1. PBK modelling .....  | 21 |
| 6.4.2. Transcriptomics.....   | 22 |
| 6.4.3. High-throughput assays evaluating effects on targets related to human safety (e.g. <i>in vitro</i> pharmacological profiling, ToxCast, cell stress assays) ..... | 23 |
| 6.5. Risk assessment output .....   | 24 |
| 7. DISCUSSION.....  | 25 |
| 7.1. Ensuring rigour in toxicokinetic and metabolite predictions.....   | 25 |
| 7.2. Determining whether biological coverage is broad enough.....   | 27 |
| 7.3. Being explicit about the level of confidence in the assessment.....  | 29 |
| 7.4. Agreed standards for using tools and reporting data.....   | 30 |
| 7.5. Ability to distinguish adaptation from adversity .....   | 32 |
| 7.6. Updated risk assessment workflow .....   | 33 |
| 7.7. More case studies .....  | 35 |
| 8. CONCLUSIONS.....   | 36 |
| 9. ACKNOWLEDGEMENTS.....  | 37 |
| 10. REFERENCES.....   | 37 |

## 1. SELECTED ABBREVIATIONS AND DEFINITIONS

|                   |  |
|-------------------|--|
| AOP               | Adverse outcome pathway  |
| AUC               | Area under the curve   |
| BER               | Bioactivity:exposure ratio   |
| BMD <sub>10</sub> | Benchmark dose for a 10% response of a predetermined endpoint  |
| C <sub>avg</sub>  | Average blood concentration over time  |
| C <sub>max</sub>  | Maximum predicted plasma concentration   |
| C <sub>ss</sub>   | Concentration at steady state  |
| ER                | Endoplasmic reticulum  |
| GIVIMP            | Good <i>in vitro</i> method practice   |
| GLP               | Good laboratory practice   |
| HepG2             | Human hepatoma cell line   |
| HepaRG            | Human bipotent progenitor cell line (capable of differentiating toward biliary-like and hepatocyte-like cells) |
| HCS               | High-content screening   |
| HTS               | High-throughput screening  |
| HTTr              | High-throughput transcriptomics  |
| ICCR              | International Cooperation on Cosmetics Regulation  |
| iTTC              | Internal threshold of toxicological concern  |
| LRSS              | Long Range Science Strategy  |
| MCF-7             | Michigan Cancer Foundation-7 (human breast cancer) cell line   |
| MIE               | Molecular initiating event   |
| MPS               | Micro-physiological systems  |
| NAM               | New approach methodology   |
| NGRA              | Next generation risk assessment  |
| NOTEL             | No-observed-transcriptional-effect level   |
| OECD              | Organisation for Economic Co-operation and Development   |
| PBK               | Physiologically-based kinetic  |
| PoD               | Point of departure   |
| QSAR              | Quantitative structure-activity relationship   |
| RhE               | Reconstructed human epidermis  |
| SEURAT-1          | Safety Evaluation Ultimately Replacing Animal Testing-1  |

## 2. BACKGROUND

In 2017 and 2018 respectively, the International Cooperation on Cosmetics Regulation (ICCR) Integrated Strategies for Safety Assessment of Cosmetic Ingredients Joint Working Group (JWG) produced reports describing the principles underpinning the use of new approach methodologies (NAMs) in the risk assessment of cosmetic ingredients, and guidance on how NAMs can be incorporated into next generation risk assessment (NGRA) (Amaral *et al.*, 2017; Amaral *et al.*, 2018). The nine ICCR principles were later published as a peer-reviewed paper (Dent *et al.*, 2018). The JWG concluded that many of the tools available to perform NGRA are already either in use in cosmetic safety assessment or are mature technologies with likely utility.

In 2019, the JWG organized a scientific workshop to investigate how the nine ICCR principles are currently being applied to NGRA case studies being performed in different organizations and to explore how application of these principles can aid safety decision in risk assessments which use NAMs.

Three case studies were presented at the workshop. Each case study described how non-animal approaches have been used to complete an exposure-led risk assessment and covered a variety of health effects relevant to cosmetics. Workshop participants discussed how application of the nine ICCR principles can underpin the use of new approaches in safety decision making. Gaps that may prevent decision making were also identified and discussed.

The JWG invited several internationally renowned experts in cosmetics safety assessment and the use of NAMs, and the agenda was as follows:

|   |
|---|
| <b>Agenda item</b>  |
| Welcome and background (introduction to the ICCR and the work of the JWG)                       |
| Workshop objectives   |
| Keynote: Using New Approach Methods for Decision Making   |
| Case Study 1: Phenoxyethanol in a body lotion   |
| Case Study 2: Coumarin in face cream and body lotion  |
| Case Study 3: Local toxicity of quasi-drugs   |
| Breakout Session 1 (each participant selected one case study group to join for discussion)      |
| Break   |
| Breakout Session 2 (each participant selected a second case study group to join for discussion) |
| Feedback in plenary (discuss commonalities/gaps across case studies)                            |
| Discussion in Plenary   |

### 3. PURPOSE

The objectives of the workshop were:

- To explore whether NGRA for cosmetic ingredients is as protective of human health as is traditional (animal-based) safety assessment
- To review some examples of NGRA for cosmetic ingredients, agree on the common features that were working well, and identify data gaps
- To agree on the next steps needed to make NGRA a day-to-day reality for the safety assessment of cosmetic ingredients

This ICCR report describes the outcome of this workshop, which was also published<sup>1</sup> in the journal *Regulatory Toxicology and Pharmacology*. Note that the main focus of this report are the systemic case studies, because NGRA for systemic safety represents the greatest challenge in the application of non-animal approaches.

---

<sup>1</sup> <https://dx.doi.org/10.1016/j.yrtph.2021.105026>

## 4. INTRODUCTION

Ethical principles and legal requirements have resulted in an increasing global interest in bringing safe cosmetic products to the market without animal testing. There is also increasing interest in applying non-animal approaches more broadly than cosmetics, to industrial chemicals and environmental contaminants (Thomas *et al.*, 2019). One vision for integrating new types of data into safety decision-making is termed next generation risk assessment (US EPA, 2014). NGRA offers a possible way forward for animal-free safety decision-making, and in 2018 the ICCR published nine principles for the NGRA of cosmetic ingredients (Dent *et al.*, 2018). Figure 1 summarizes these nine principles, of which four are overriding: (1) the overall goal is a human safety risk assessment that is (2) exposure-led, (3) hypothesis-driven and (4) designed to prevent harm. Additional principles describe how an NGRA should be conducted: following an appropriate appraisal of all existing information, using a tiered and iterative approach, and using robust and relevant methods and strategies. Finally, the principles outline that the assessment should be documented transparently, being explicit about the logic of the approach and sources of uncertainty.

- 
- 4 Main overriding principles:**
    - » The overall goal is a human safety risk assessment
    - » The assessment is exposure led
    - » The assessment is hypothesis driven
    - » The assessment is designed to prevent harm
  - 3 Principles describe how a NGRA should be conducted:**
    - » Following an appropriate appraisal of existing information
    - » Using a tiered and iterative approach
    - » Using robust and relevant methods and strategies
  - 2 Principles for documenting NGRA:**
    - » Sources of uncertainty should be characterized and documented
    - » The logic of the approach should be transparently and documented

**Figure 1:** The nine ICCR Principles of Next Generation Risk Assessment (NGRA) of cosmetic ingredients (Dent *et al.*, 2018).

## 5. THE LEVEL OF CONSUMER PROTECTION OFFERED BY NGRA

The traditional approach to ensuring the safety of cosmetic ingredients has been to evaluate the compounds in animal-based toxicity tests. The range of tests have evolved over time, but typically involve the evaluation of local and systemic effects for a chemical and include clinical observations as well as gross and histopathological endpoints. When performed in dose response, the animal-based tests intend to identify both the dose at which adverse effects begin to occur as well as the dose range below which no detectable adverse effects are observed. Historically, the information derived from these studies form the basis for a risk assessment and protection of the consumer.

It is challenging to fully assess the level of consumer protection afforded by the traditional animal testing approaches. Studies that have evaluated the animal-to-human concordance during pre-clinical and clinical safety testing of pharmaceuticals (Ackley *et al.*, 2019; Monticello *et al.*, 2017) indicate that most animal-based toxicity tests show limited positive predictive value for specific toxicological endpoints, but they have relatively high negative predictive value (Monticello *et al.*, 2017). In other words, the traditional animal-based testing paradigm is good at identifying the absence of potential human toxicity, but they are not necessarily predictive of specific health hazards in humans.

With the overall goal of a human safety risk assessment, the lessons learned from evaluating the animal-to-human concordance for pharmaceuticals may have relevance for designing the NGRA for cosmetic ingredients. Developing *in vitro* assays, *in silico* (computational) models, or batteries of assays that predict specific endpoints or responses in animals and validating the methods to demonstrate accuracy and reproducibility in the results is not feasible to replace the complex, multi-endpoint animal studies for systemic toxicity. In addition, given the low positive predictive value of the animal studies for human responses (Monticello *et al.*, 2017), the potential relevance of specific endpoints from animal studies may be limited. Instead, an alternative strategy for the NGRA would be to develop and employ *in vitro* and *in silico* (computational) methods in a manner that aims to ensure the absence of potential human toxicity (*i.e.*, consumer protection), while still being predictive

of specific human hazards where the linkage between the molecular initiating event and the adverse outcome is well established. Recent work has demonstrated that biological activity across a diverse battery of *in vitro* assays can be used as a conservative estimate of a quantitative PoD across a broad range of traditional animal toxicity study designs including subchronic, chronic, developmental, and reproductive studies (Paul Friedman *et al.*, 2020). Other studies exploring computational methods have linked *in vitro* assays for specific molecular initiating events such as estrogen receptor bioactivity and adverse outcomes (Browne *et al.*, 2015). The integration of these two approaches form the foundation of what can be done in an NGRA.

## **6. COMMON FEATURES IN NGRA FOR COSMETIC-RELEVANT CASE STUDIES**

Case studies play a pivotal role in assessing and applying new approaches (Andersen *et al.*, 2011), and several are being conducted to assess the feasibility of using NGRA to make safety decisions for potential systemic effects of cosmetic ingredients. Focus can be on the use of NGRA for local effects, especially where some national regulations require animal testing. One example is the Japanese National Institute of Health Sciences which is considering how to build a framework for the assessment of skin irritation of quasi-drugs. This framework was discussed at the workshop, and opportunities for integrating the ICCR Principles within it were considered (**see Section 6.1**).

The applicability of NGRA is also being explored for systemic effects, *e.g.* to see if non-animal approaches are sufficient to assure the systemic safety of coumarin in leave on cosmetic products (**Section 6.2.1**) (Baltazar *et al.*, 2020). The Cosmetics Europe Long Range Science Strategy (LRSS) is aimed at applying the SEURAT-1 risk assessment workflow (Berggren *et al.*, 2017) to several cosmetics-relevant case studies (Desprez *et al.*, 2018). One of these case studies is to assess the applicability of NGRA in the safety assessment of the commonly used cosmetics preservative phenoxyethanol at a use level of 1% in a body lotion (**Section 6.2.2**).



Both systemic toxicity case studies (coumarin and phenoxyethanol) were explored and challenged by workshop attendees, with an emphasis on the ability of the case study to support robust safety decision-making. In the NGRA for systemic toxicity of both, coumarin and phenoxyethanol, there are several common points that can be categorised as: 1. Application of the ICCR principles; 2. Use of common tools and approaches; 3. Risk assessment output.

### **6.1. Application of ICCR Principles to evaluation of local toxicity**

Performing a safety assessment for local effects such as skin irritation is very different from addressing more complex systemic effects. In many regions, risk assessments for the local toxicity of dermally applied ingredients can be completed without performing any new animal experiments. This is done using a variety of techniques, including consideration of the physicochemical properties of the ingredient, its concentration in the formulation, application of *in silico* tools, read-across, or benchmarking the irritancy of the ingredient or formulation using *in vitro* tests (Macfarlane *et al.*, 2009). However, in some regions regulations still require the generation of animal test data to assess the local toxicity of dermally applied products and/or their ingredients. For example, application for a new dermal quasi-drug in Japan requires the generation of primary and cumulative irritation data over 24-hours by animal testing (MHLW, 2011). Such data are requested based on the concern that substances predicted *in vitro* as being non-irritating after short exposure (4hrs) (OECD 404 or OECD 439), may become irritating *in vivo* after a longer application period.

The ICCR Principles described in Figure 1 offer a rationale for developing a workflow, particularly in this situation of skin irritancy to arrive at a risk assessment decision for a novel cosmetic or quasi-drug that does not involve the generation of new animal data. An exposure-led approach would consider the concentration of the ingredient in the formulation, its physicochemical properties and other existing data to determine whether at that level the ingredient is likely to increase the irritation potential of the formulation. In case there are insufficient data to make conclusions about the irritation potential of the ingredient, the *in vitro* tests selected to be performed should be capable of

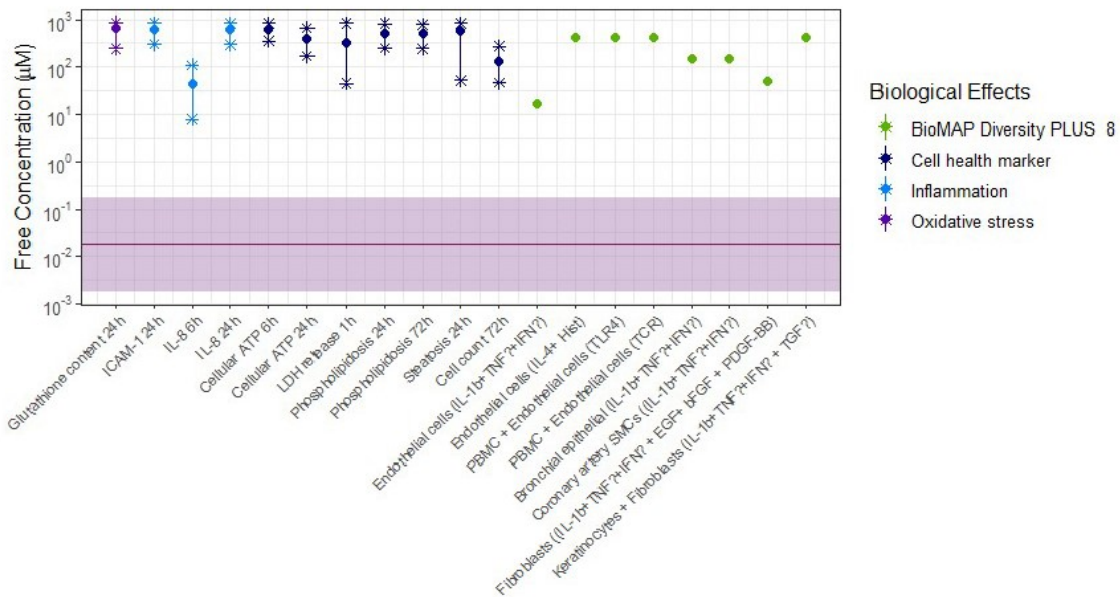
determining if there is a high potential for skin irritation following extended exposures to consumers. The reconstructed human epidermis test method for *in vitro* skin irritation (OECD 439) describes a protocol for identifying irritant substances. It uses reconstructed human epidermis (RhE), which closely mimics the biochemical and physiological properties of the upper parts of the human skin, and therefore also meets the ICCR principle of human relevance. In cases where the test shows no irritation potential, it is considered highly unlikely that longer exposures in humans would lead to serious skin irritation. Even in situations where effects indicating an irritation potential are seen, this information could be used alongside knowledge of the level of the ingredient in the formulation and product use information to determine whether a safety risk exists for consumers, benchmarking against other formulations as appropriate. Overall, workshop participants agreed that the ICCR principles offer opportunities to prevent unnecessary animal tests for local effects.

## **6.2. Systemic toxicity case studies**

### **6.2.1. Overview of coumarin case study**

The case study of coumarin at 0.5 % in a shampoo and face cream discussed at the workshop, illustrated the usefulness of NGRA for systemic toxicity. The approach used was similar to a previously described case study for the same ingredient used at 0.1% in face cream and body lotion (Baltazar *et al.*, 2020), where full details of the assays and underpinning data can be found. In-line with the nine ICCR principles, this case study involved first performing a literature search (excluding any available *in vivo* animal data). Next, *in silico* tools were used to identify potential pathways of toxicological concern and to formulate hypotheses for adequate *in chemico* and *in vitro* testing. Internal exposures to coumarin were estimated using a physiologically based kinetic (PBK) model; the maximum predicted plasma concentration reached ( $C_{max}$ ) following dermal application was used as the internal dose metric. The case study authors indicated that the PBK model output showed a high level of concordance with human clinical data as reported elsewhere (Moxon *et al.*, 2020). The  $C_{max}$  value was compared with PoDs generated from a battery of *in chemico* and *in vitro* NAMs. These NAMs were

selected to cover a variety of biological effects such as receptor-mediated and immunomodulatory effects (BioMap Diversity 8 (Houck *et al.*, 2009)), and general bioactivity (an *in vitro* cell stress panel (Hatherell *et al.*, 2020) and high-throughput transcriptomics (HTTr)). The PoDs from the *in vitro* assays (expressed as free concentration ( $\mu\text{M}$ ) in culture medium) were plotted against the predicted human *in vivo* exposure (expressed as free concentration ( $\mu\text{M}$ ) in plasma) to calculate a bioactivity:exposure ratio (BER) with associated uncertainty, as illustrated for the BioMap and cell stress panel data in Figure 2. The product with the highest  $C_{\text{max}}$  value was face cream (Moxon *et al.*, 2020); this value was lower than all PoDs with BER values greater than 100, the assessment factor used in traditional (animal-based) safety assessment. Key work that was identified to strengthen the case study was the generation of further experimental data to understand the influence of skin and liver metabolism. A critical unresolved question for this case study at the time of the workshop was whether the metabolites formed *in vivo* following consumer use of the cosmetic would be present in the *in vitro* systems used to provide the PoDs, and the influence this would have on the safety decision made. A challenge identified by this case study was how to assure the quality and robustness of non-standard (non-test guideline and/or non-GLP) data and to characterize uncertainty to allow informed decision-making.



**Figure 2:** Points of departure (PoDs) plotted against predicted plasma exposure for coumarin case study (face cream exposure). PoDs expressed as free concentration in assay/plasma. Red line = Plasma  $C_{max}$  free ( $\mu\text{M}$ ); pink band = uncertainty band in plasma level prediction. BERs calculated to be between 2500 and 38000.

### 6.2.2. Overview of phenoxyethanol case study

The overall approaches and data used in this case study were very similar to the published coumarin case study (Baltazar *et al.*, 2020). The overarching hypothesis was that systemic exposure to phenoxyethanol present at 1% in body lotions will not cause adverse health effects in consumers. An adverse effect was defined as a biochemical, morphological or physiological change that either singly or in combination negatively affects the consumer or reduces their ability to respond to an additional environmental challenge (Lewis *et al.*, 2002). The consumer was defined as any individual who is likely to use a body lotion placed onto the market, including high-end (90<sup>th</sup> percentile) users. This case study used the SEURAT-1 *ab initio* workflow as a basis (Berggren *et al.*, 2017) and was conducted in a tiered and iterative manner. As per this framework, exposure-based waiving and read-across were both considered but could not be used due to the high exposure level and lack of structurally similar

substances with systemic toxicology data. PBK modelling was based on a previously published model (Troutman *et al.*, 2015) with the exclusion of data generated in animals. The case study authors indicated that the PBK model was independently verified using human data, and there was a high confidence in the model output. Identification of significant metabolites was performed following a tiered approach, whereby *in silico* predictions were followed up with generation of *in vitro* metabolism data in primary human hepatocytes. This tiered approach identified the need to focus the risk assessment on both phenoxyethanol and its major metabolite phenoxyacetic acid. The internal (plasma) exposure to both these molecules was therefore calculated using PBK modelling. In addition to  $C_{max}$  calculations, the case study authors attempted to include a time component to the safety assessment by calculating the average blood concentration over time ( $C_{avg}$ , calculated by dividing the steady state area under the blood concentration/time curve by 24, the duration in hours of the assays used to provide the point of departure). For phenoxyethanol the  $C_{max}$  and  $C_{avg}$  values were 3.2 and 0.18  $\mu\text{M}$  respectively, and for phenoxyacetic acid they were 2.7 and 1.7  $\mu\text{M}$  respectively.

Published (non-animal) data were evaluated and *in silico* screening for bioactivity was performed for both the parent molecule and its main metabolite in an effort to identify modes of action that could form a safety concern. Data on both phenoxyethanol and phenoxyacetic acid from ToxCast and the PubChem database were evaluated. This showed that both substances exhibited very little biological activity in the available assays. To identify a PoD to use in the risk assessment, HTTr data were generated in 3 cell lines (HepaRG, HepG2 and MCF-7 cells) and a no-observed-transcriptional-effect level (NOTEL) was calculated (**see Section 6.4.2**). The exposure values, PoDs and resulting bioactivity exposure ratios are presented in Table 1.

**Table 1:** Bioactivity:exposure ratios (BERs) calculated for phenoxyethanol

|        | NOTEL (BMD <sub>10</sub> ) for lowest pathway altered (µM) | BER based on C <sub>max</sub> of 3.2 µM | BER based on C <sub>avg</sub> of 0.18 µM |
|--------|--|---|--|
| HepG2  | 442  | 138                                     | 2456                                     |
| HepaRG | 60   | 19                                      | 333                                      |
| MCF-7  | 150  | 47                                      | 833                                      |

A major limitation of the data available at the time of the workshop was the lack of HTTr data on phenoxyacetic acid, or confirmation that this metabolite was generated in the bioactivity assays performed. It was therefore not possible arrive at a conclusion for this case study. The challenges associated with this case study were definition of a BER that would be considered low risk, and characterising the uncertainty, especially concerning the range of biological activities and hence potential toxicities that were covered by the limited set of cell lines (three in total) used in the HTTr experiments.

### **6.3. Application of ICCR Principles for the evaluation of systemic toxicity**

#### **6.3.1. The overall goal is a human safety risk assessment**

This ICCR principle states that “[t]he safety assessment should enable a decision to be made on the safety of the ingredient/product to humans, not be designed as a prescriptive or definitive battery of tests to replicate the results of animal studies” (Dent *et al.*, 2018). This principle was considered to be very evident in the systemic case studies discussed. Both case studies used *in vitro* models and tools that were selected because they had a basis in human biology, and used exposure information to arrive at a decision thought to be relevant to the population of interest. Crucially, neither case study used any pre-existing animal data (thus emulating the ‘animal free’ development of a novel molecule) and did not attempt to predict/interpret adverse effects in animals.

### 6.3.2. *The assessment is exposure led*

*“In an exposure-driven paradigm exposure estimates will define the degree of hazard data needs and guide further data generation”* (Dent *et al.*, 2018). In both examples, a tiered approach to exposure and the identification of major metabolites was taken. This led the phenoxyethanol case study authors to focus their attention not only on the parent molecule but also on the major stable metabolite phenoxyacetic acid, and to the conclusion that further data needed to be generated to complete the safety assessment. Although the same considerations were applied in the coumarin case study, the large BERs obtained at the end of the assessment (well over 100) led the case study authors to conclude that a decision could be made without further data generation. In addition, the coumarin safety assessment was based on free concentration values, which provides a further level of refinement to the safety assessment. Internal dose metrics used to perform a safety assessment can include the concentration at steady state ( $C_{ss}$ ), area under the curve (AUC) and  $C_{max}$  (Wambaugh *et al.*, 2018). Both case studies used  $C_{max}$  as a key metric, which although thought to provide a conservative comparison, does not consider exposure over time. The phenoxyethanol case study therefore also used  $C_{avg}$  in an attempt to provide a more relevant/less conservative comparison with the static *in vitro* situation. A further refinement in approach that would allow for a more robust and realistic comparison between *in vitro* and *in vivo* exposures could involve measuring the test chemical in the *in vitro* system over time, thus enabling a comparison of AUC *in vitro* with the AUC predicted *in vivo*.

### 6.3.3. *The assessment is hypothesis drive*

The hypotheses tested in NGRA may be specific to a mode of action (e.g. *“At relevant exposures, chemical X perturbs the p53 pathway which results in increased cancer risk in consumers”*) or may be general, e.g. *“At relevant exposures the biological activity of Chemical X is insufficient to cause adverse effects in consumers”*. This latter hypothesis may be adequate where there is a wide margin between any relevant *in vitro* activity and human exposures (Dent *et al.*, 2018) and was the basis for these case

studies. The overarching hypothesis was therefore that the biological coverage provided by the *in vitro* assays was sufficient to provide assurance that at human-relevant exposures there was no biological activity that could lead to an adverse health effect. In addition, each case study considered whether there were specific modes of action that needed to be addressed to arrive at a safety decision. For example, phenoxyethanol inhibits bacterial malate dehydrogenase (Gilbert, Beveridge, & Crone, 1977a, 1977b). If the human form of this enzyme were inhibited at consumer-relevant concentrations this could result in adverse effects on oxidative phosphorylation. This concern was ultimately addressed (following the workshop) using published *in vitro* data which concluded that phenoxyethanol has no adverse effects on respiration in human cells (Hatherell *et al.*, 2020).

#### 6.3.4. *The assessment is designed to prevent harm*

*“Where no biological activity is predicted to occur at human-relevant exposures there can be no adversity. However, many NAMs can identify biological effects with great sensitivity, meaning that where biological activity is predicted at consumer-relevant exposures, tools and approaches distinguish between an adaptive and an adverse response are needed”* (Dent *et al.*, 2018). Both systemic case studies followed a ‘protective not predictive’ approach. In the coumarin case study, there was no *in vitro* biological activity at concentrations relevant to consumer exposures. The assessment was therefore protective of consumers but not predictive of adverse effects (pathology) that could occur at higher exposures. For the phenoxyethanol case study, insufficient data were available to illustrate this principle due to missing critical data for the major metabolite phenoxyacetic acid, which the case study authors planned to address by generating further data. An important point to note about both assessments is that although they were ‘Designed to Prevent Harm’, neither assessment attempted to differentiate between biological activity and adversity. Where biological activity cannot be ruled out at human-relevant exposures (*i.e.*, the BER is small), there would be a need to understand whether this activity could result in an adverse health effect using more complex and physiologically relevant *in vitro* or *in silico* approaches. The case study authors attempted to make



a safety decision within their given context, considering the capabilities/limitations of the toolbox and based on biological activity alone. It should be noted that this assumes there is confidence that the models/assays used adequately cover the biological space (*i.e.*, the modes of action that could lead to adverse effects), and that the concentration range of testing is broad enough to characterize when bioactivity occurs.

However, the lack of tools currently available to distinguish between biological activity and adversity remains a major limitation in NGRA.

#### *6.3.5. Conduct an appropriate appraisal of existing information*

This principle reinforces the importance of ensuring that all available relevant knowledge and information is used to shape the scope and direction of the assessment. This principle was not fully applied because much existing data (*e.g.* that derived from history of use or previous *in vivo* experiments) was excluded to simulate the safety assessment of a novel molecule.

#### *6.3.6. Use a tiered and iterative approach*

*“The total amount of resources allocated to any risk assessment should be no less and no more than that required to provide adequate precision, to reach a conclusion, and to make a decision.”* (Dent *et al.*, 2018). Adhering to this principle means that the data are generated in an efficient way. For example, before embarking on an exposure-led NGRA it would be sensible to ensure there are no major safety liabilities that would be difficult to overcome, such as genetic toxicity or severe acute toxicity (*e.g.* using quantitative structure-activity relationships (QSARs) and established *in vitro* tests). Both systemic case studies discussed at the workshop followed a tiered approach in which successive layers of refinement were applied to increase confidence in the outcome of the assessment. Although some uncertainties remained, data generation was stopped in the coumarin case study once the authors determined they could reach a decision. For the phenoxyethanol case study the need for

further data generation was identified to arrive at a robust conclusion based on the high predicted exposures to the major acid metabolite. Both case studies therefore exemplified this principle.

#### 6.3.7. *Use robust and relevant methods and strategies*

The use of relevant and robust methods is a basic need to enable confident decision-making. This requires use of tools and approaches that are adequate for informed risk assessments by providing data of sufficient quality to use in decision-making. This principle states that *“[i]n determining the usefulness of a method, the applicability domain and limitations of the method need to be well understood and documented, so that the methods can be applied appropriately. The relevance of the method for the specific purpose also needs to be considered and justified”* (Dent *et al.*, 2018). The exposure approaches used were in general considered reliable and relevant to use in safety decision making. The bioactivity measurements were based on an understanding of modes of action that could lead to adverse effects in humans (*e.g.* cellular stress), *in vitro* pharmacology screening that has a track record of reducing safety-related drug attrition, and whole genome transcriptomics for broad biological coverage. However, a detailed review of the response of all the tools and analysis techniques used in these case studies was outside the scope of the workshop. Although individually the tools used were robust, more work is needed to assess their relevance when used as a panel for cosmetics safety assessment, and there was concern that all possible activities of toxicological relevance were not captured (**see Discussion Section 7.2**).

#### 6.3.8. *Characterize and document sources of uncertainty*

*“All sources of uncertainties should be identified and characterised to provide transparency for the decision-making process, ideally leading to a future where default ‘uncertainty factors’ are redundant”* (Dent *et al.*, 2018). The coumarin case study identified some critical areas of uncertainty, but, given the large BERs the case study authors did not consider that these prevented a safety decision being made. In contrast, the key area of uncertainty in the phenoxyethanol case study (the bioactivity of the major acid metabolite) prevented the case study authors from making a safety assessment

conclusion. Although both case studies considered key areas of uncertainty, these were not documented in a systematic way for workshop participants to evaluate.

#### 6.3.9. Document the logic of the approach transparently

*“All data used, assumptions, methodology and software should be clearly documented and be available for independent review.”* The safety assessments were presented in a logical and reasoned manner, although the nature of the workshop was that a thorough review of a documented assessment was not conducted. It should however be noted that, following the workshop, a modified version of the coumarin case study has been documented in the peer-reviewed literature (Baltazar *et al.*, 2020). Furthermore, the phenoxyethanol case study has been published as an OECD integrated approach to testing and assessment (OECD, 2021a)<sup>2</sup>. Therefore, the logic of the strategy underlying each case study is now available for broader scrutiny.

#### 6.4. Common tools and approaches in the evaluation of systemic toxicity

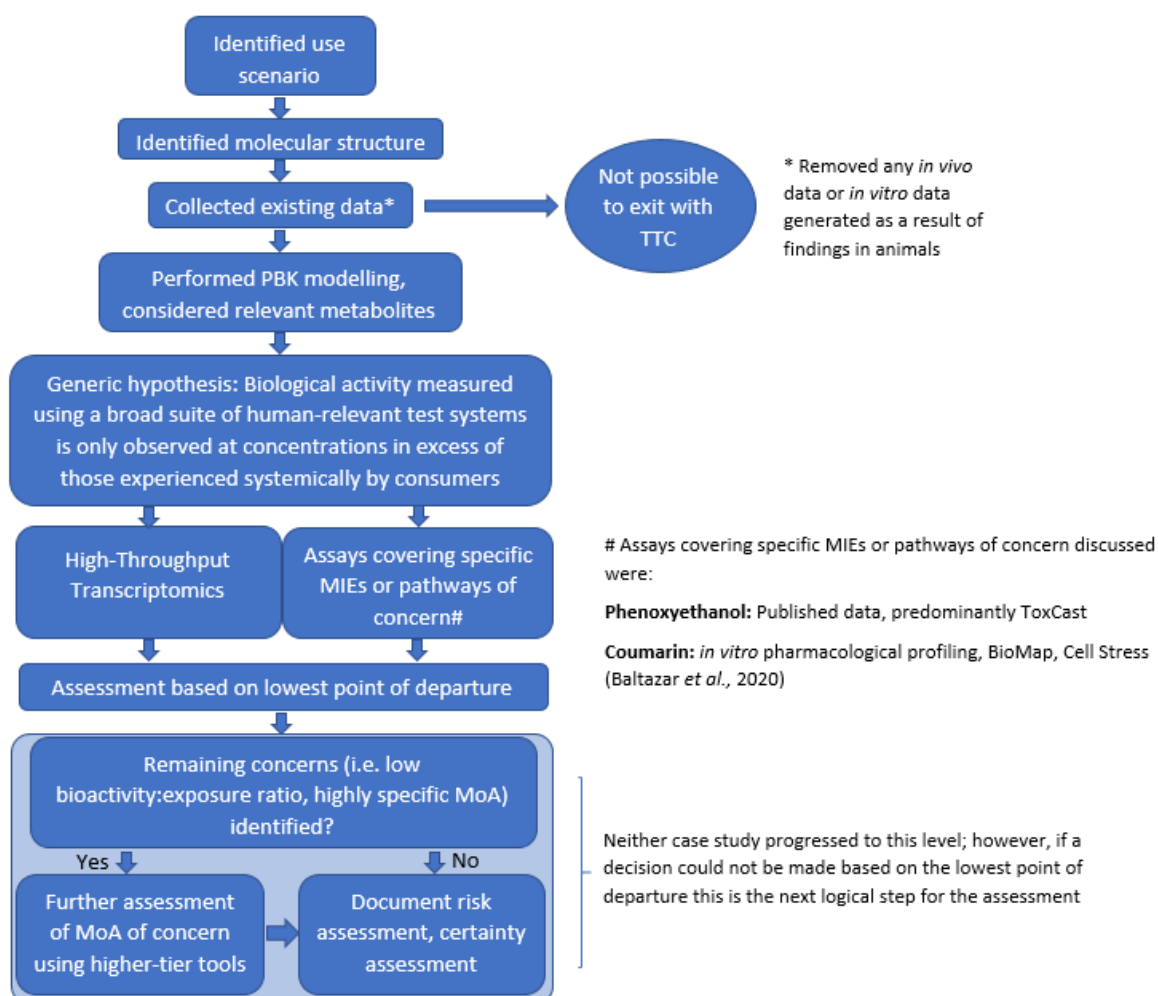
Although differences between the systemic toxicity case studies existed, the overall objective of both relied on the generation of a suite of *in vitro* bioactivity data with broad biological coverage. These data were used alongside specific assays that should provide information whether specific key molecular initiating events (MIEs) (Allen *et al.*, 2014) that are known to be associated with particular adverse events and are not triggered at relevant exposures. This *in vitro* bioactivity was compared with the *in silico* (PBK model) predicted internal exposure of the ingredient using *in vitro* to *in vivo* extrapolation. The margin between the *in vitro* bioactivity and predicted internal exposure concentrations was used to determine whether *in vivo* systemic bioactivity as an indicator of adverse effects, was likely to occur or not. The common tools and overall approach used to illustrate the application and integration of NAMs for the evaluation of systemic effects for two cosmetic

---

2

[https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/CBC/MONO\(2021\)35&docLanguage=En](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/CBC/MONO(2021)35&docLanguage=En)

ingredients across both case studies are described in Figure 3; details for each of the workflow components are further outlined in the sections below.



**Figure 3:** Common components in the application of the SEURAT-1 risk assessment workflow to two case study safety assessments for systemic effects.

For both case studies, ToxCast data were available. These data were used in the phenoxyethanol case study, but not considered in the coumarin case because the case study authors took the decision to emulate the development of a new cosmetic ingredient, where no ToxCast data would be available. Instead, coumarin was evaluated using *in vitro* pharmacological profiling was performed (using the SafetyScreen44™ panel (Bowes *et al.*, 2012)) and the BioMap Diversity 8 Panel to characterize effects relevant to safety pharmacology and inflammatory disease (Houck *et al.*, 2009). In addition, data on

the potential for coumarin to cause cellular stress were investigated *in vitro*. It should be noted that, although not available at the time of the workshop, both *in vitro* pharmacological profiling and cellular stress panel data were also generated for phenoxyethanol, and parent molecule and metabolite kinetics were defined in the cell lines used to establish the NOTEL to better inform the safety assessment. Alongside PBK modelling, the critical *in vitro* tools to assess bioactivity for both case studies (transcriptomics complemented with more specific assays evaluating cellular targets known to give alerts for human safety) are briefly described below.

#### 6.4.1. PBK modelling

A NGRA for systemic effects is likely to rely heavily on PBK modelling to predict systemic exposure and to guide the assessment, *e.g.* by predicting organs or systems that are most highly exposed to the parent molecule or a metabolite. The PBK models used in the case studies relied on *in vitro* inputs including either published or newly generated skin penetration data and hepatocyte clearance data. Predicting the systemic levels of the parent molecule and metabolites is critical to ensure that a robust safety decision can be made. Making such predictions and verifying them in the absence of any *in vivo* data is a major challenge in NGRA. Both systemic case studies attempted to use a tiered (*in silico* first) approach, including in the identification of key metabolites, and generated or used existing *in vitro* data to confirm or refute the predictions. One limitation in both systemic case studies was that the topic of the case study (the chemical of interest) was likely to be represented in the dataset used to inform the knowledge-based system that was applied to predict metabolic products (Meteor Nexus version 3.1.0 (Lhasa Ltd.)). It will therefore be helpful to conduct further case studies on novel chemistries, to assess the performance as well as the efficiency of the approach used, and to better address aspects of uncertainty. It was also highlighted that beyond more case studies, more *in silico* efforts are needed towards predicting dermal penetration and metabolism.

#### 6.4.2. Transcriptomics

Techniques such as transcriptomics (a description of all the genes expressed within a cell or tissue), is likely to have an integral role in supporting NGRA, and for both systemic toxicity case studies, bioactivity data with the broadest biological coverage were whole genome transcriptomics. Transcriptomics data are central to the NGRA approach for cosmetic ingredients and these data may be integrated in several areas in the risk assessment. For example, gene expression profiling may be beneficial in informing read-across, to allow grouping of chemicals that have the same mode of action (De Abrew *et al.*, 2016, 2019). In addition, transcriptomics data may have value in helping to understand the biological pathways that a chemical may affect developing or testing mode of action hypotheses (Catlett *et al.*, 2013; S. Labib *et al.*, 2016). However, the way that transcriptomics data were applied to both of these case studies was using the concept of a no-observed-transcriptional effect level or NOTEL (Lobenhofer *et al.*, 2004). A NOTEL is the test concentration that does not result in any differential gene expression compared with control, or from a systems biology point of view does not result in any meaningful changes. In other words, isolated changes in individual genes may not represent a meaningful biological difference, whereas changes in the expression of several genes within a pathway may be biologically meaningful (Farmahin *et al.*, 2017). Although the strength of using a NOTEL in risk assessment is the potential to cover the whole genome in multiple cell types, a key area where consensus is needed is which cells are needed to ensure broad biological coverage (**see Discussion Section 7.2**). One consideration with respect to transcriptomics data is that for some toxicities changes in gene expression may not be the most sensitive effect. It is therefore important that transcriptomics data be used alongside complementary assays to provide enough information on the modes of action for which the concordance between gene expression changes and adverse effects is not well understood. Finally, a limitation of concern to several workshop participants is that changes in gene expression do not always correlate with changes in protein expression and questioned whether a NOTEL is a relevant and useful point of departure for risk assessment. However, other participants pointed to the increasing scientific evidence of the concordance between changes in gene

expression and pathological change (Clewell *et al.*, 2011; Sarah Labib *et al.*, 2016; Qutob *et al.*, 2018; Thomas *et al.*, 2007, 2011), and highlighted further examples published after the workshop (Farmahin *et al.*, 2019; Gwinn *et al.*, 2020). Overall, the engaging discussions on this topic highlighted the need to ensure agreement on standards for using these novel tools and reporting the data to increase confidence they can be applied in a robust, consistent and useful way.

#### 6.4.3. High-throughput assays evaluating effects on targets related to human safety (e.g. *in vitro* pharmacological profiling, ToxCast, cell stress assays)

The phenoxyethanol case study relied heavily on ToxCast assays for both the parent molecule and, to a lesser extent (due to testing in a limited number of assays), its major metabolite phenoxyacetic acid. The original vision of ToxCast was to provide a general bioactivity screen using existing assays and covering mechanisms and pathways including signal transduction, receptor-mediated effects, apoptosis, oxidative stress, DNA repair and cell cycle (Dix *et al.*, 2007). These mechanisms and pathways were selected for their relevance to high hazard substances such as reproductive and developmental toxicants and carcinogens. Although the ToxCast programme was conceived as a prioritization tool and therefore has limitations in terms of its coverage (including limitations in metabolism), the PoDs it provided do appear to be conservative and protective of various adverse effects when compared with *in vivo* PoDs (Paul Friedman *et al.*, 2020). For a novel chemical substance, ToxCast data will not be available. However, similar approaches assessing targets associated with known safety liabilities such as functional assays evaluating key signalling pathways may form an important part of some assessments. For example, although the ToxCast data on coumarin were not included in the overall evaluation, data investigating the ability of the parent molecule to cause changes in biomarkers associated with both inflammatory disease (Houck *et al.*, 2009) and cellular stress as well as *in vitro* pharmacological profiling were generated using other tools.

The rationale for inclusion of a panel of cellular stress assays in the coumarin case study was that substances that do not interact with specific biological targets may still cause adverse health effects

by causing cellular stress responses. Cellular stress responses are critical in ensuring homeostatic control following exposure to xenobiotics (Simmons *et al.*, 2009), so information on whether a substance causes cellular stress at relevant exposures may form an important part of an *in vitro* toxicological evaluation. The cell stress pathways considered included oxidative stress, DNA damage, inflammation, endoplasmic reticulum (ER) stress, metal stress, heat shock, hypoxia, mitochondrial toxicity and general cellular health (Hatherell *et al.*, 2020). Data from these assays as well as more general cytotoxicity assays may be informative of the PoD for safety assessment in the absence of other more sensitive or specific biological targets.

The type of *in vitro* pharmacological profiling represented by the SafetyScreen44™ panel involves screening compounds against a range of targets (receptors, ion channels, enzymes and transporters) using a variety of assay technologies (Bowes *et al.*, 2012). This type of approach is used in early drug discovery as a screen for modalities that are responsible for safety-related attrition, especially for the primary organ systems. As a screen to be used in an animal-free risk assessment, the forty-four targets that form this battery are arguably too narrow. Others have sought to extend this battery to give more coverage (Lynch *et al.*, 2017), but more work is needed to arrive at a list of targets useful for the safety evaluation of cosmetic ingredients (**see Discussion Section 7.2**). The panel of targets may be tailored *e.g.* to cover *in silico* alerts and modes of action identified from other *in vitro* studies (*e.g.* transcriptomics), and therefore used in a tiered manner to help develop mode of action hypotheses for a given chemical exposure. Where hits are obtained, these should be followed up to understand the dose-response and to obtain a PoD for use in the risk assessment.

#### **6.5. Risk assessment output**

The overall objective of the systemic case studies was to explore how NAMs can be used to inform consumer exposure and upstream molecular events for safety decision-making, rather than to predict what adverse health effects may occur at higher exposures. The key assumption was therefore that a large BER indicates that *in vivo* systemic bioactivity, and therefore adverse effects, are unlikely. Where



margins are insufficient or overlap more mechanistic understanding is needed to be able to differentiate between activity and adversity, however this additional tier was not a feature of either systemic case study presented at the workshop. Critical questions for these case studies are therefore, “what BER is appropriate using *in vitro* bioactivity data?”, and whether the toolbox used had sufficient biological coverage to arrive at a confident safety decision. Several ideas were discussed to address these issues, and the results of these discussions are outlined below.

## **7. DISCUSSION**

Although animal-free safety assessment via the application of NGRA has shown significant progress in the cosmetics sector in recent years, it is not yet used for regulatory decision-making. There continues to be a pressing need for new tools and a fresh approach to the application of non-animal methods (Rogiers *et al.*, 2020). NAMs are being integrated, attempting to ensure breadth of biological coverage and robust exposure assessments to support safety assessment, but more work is needed to understand whether decisions based on NGRA are both protective and fit-for-purpose depending on the decision context. Because of the current uncertainty surrounding the use of NGRA for regulatory decision-making, it is necessary for those working in this area to be explicit about the strengths and limitations of each assessment.

The major issues that were identified by workshop participants that still need to be addressed to understand how to apply and to increase confidence in NGRA are listed below.

### **7.1. Ensuring rigour in toxicokinetic and metabolite predictions**

Predicting systemic exposure to both the parent molecule and relevant metabolites is a fundamental requirement for NGRA. Although PBK modelling is a mature science, its application in a completely animal-free context is still relatively new. Therefore, workshop attendees identified that a clear workflow describing how to develop a PBK model, how to capture uncertainty related to the model as

well as how to verify the output in the absence of *in vivo* data would greatly enhance confidence in the use of *in silico* approaches for the prediction of kinetic parameters and internal exposure. It should be noted that following the workshop a tiered framework to developing PBK models for NGRA in the context of dermally-applied cosmetics and use of an analogue-based approach to developing and evaluating PBK models have been published (Ellison & Wu, 2020; Moxon *et al.*, 2020), and OECD guidance is now available on characterisation, validation and reporting of PBK models for regulatory purposes without the availability of *in vivo* data (OECD, 2021b).

Determining when a predicted metabolite needs to be explicitly considered in the safety assessment is an area that requires further development. Although *in silico* tools are available to predict which metabolites may be formed, *in vitro* studies or accepted surrogates are required to confirm their presence and to quantify their formation and clearance. The greatest concern regarding metabolism is the formation of a toxic metabolite *in vivo* which is not formed in the *in vitro* tools used to characterize bioactivity. This evaluation needs to be performed in a robust and fit-for-purpose way, and relies on a fundamental understanding of the strengths and limitations of the *in silico* tools, metabolic competence of the cell models used and their relevance to the human *in vivo* situation. There are challenges in relying on assay data where only the parent molecule is dosed (even if the critical metabolite is formed), because the effects measured may be caused by either or both the parent molecule and the metabolite. Because these effects cannot be distinguished, the calculation of a separate BER for both the parent molecule and the metabolite based on the same assay data will be very conservative. In cases where a metabolite is expected to be formed in humans that is not (or is under-) represented *in vitro*, specific bioactivity assays may need to be performed for that metabolite. It is therefore critical to establish exposure thresholds whereby a metabolite is deemed to be relevant for further evaluation, potentially using approaches such as the internal threshold of toxicological concern (iTTC) (Ellison *et al.*, 2019). It is worthy of note that since the workshop, an

interim iTTC of 1  $\mu$ M has been proposed for chemicals in consumer products (Blackburn, Carr, Rose, & Selman, 2020).

## **7.2. Determining whether biological coverage is broad enough**

A limitation of both the coumarin and phenoxyethanol case studies is the uncertainty in whether the *in vitro* bioactivity assays and cell types provided comprehensive coverage for safety decision-making. The question of how to assess the breadth of biological coverage was therefore identified as key for NGRA. The goal is to ensure proper coverage of a wide variety of molecular/cellular events (to meet the human protection goal) to inform the safety decisions while being pragmatic in terms of implementation. This is particularly important when dealing with substances with no clear mode of action identified.

Transcriptomics is a means to extend biological coverage. Tools that may complement gene expression include (but are not limited to) safety pharmacology screening (for specific MIEs of concern (Bowes *et al.*, 2012; Lynch *et al.*, 2017; Whitebread *et al.*, 2016; Wu *et al.*, 2013)), cell stress assays (Hatherell *et al.*, 2020) and phenotypic profiling (Thomas *et al.*, 2019). Using these kinds of approaches allows specific targets of concern as well as more general changes to be assessed. It should be noted that the tools and approaches used in the coumarin and phenoxyethanol case studies are not the only techniques available, and other methodologies may be needed to provide safety assurance. For the tools that were used in these case studies, several observations were made.

Confidence in the use of transcriptomics data needs to be increased by striving to achieve general agreement on experimental conditions including duration of exposure, concentrations tested, cell types and analysis methods. This will ensure conditions are standardised (as appropriate), results are reproducible, and the variation range in response is understood. It is critical that data are available demonstrating that these data and analysis methods are protective of human health, using substances for which *in vivo* data exist.

Regarding the safety pharmacology screen, a pragmatic approach based on the conditions established for screening drugs' off-target effects was used, but the relevance of these targets to cosmetics is yet to be determined. Furthermore, it should be remembered that the safety pharmacology screens currently available are an evolution of those used to reduce safety-related drug attrition (Bowes *et al.*, 2012; Lynch *et al.*, 2017; Whitebread *et al.*, 2016; Wu *et al.*, 2013). This represents a very different decision context to performing an animal-free cosmetic risk assessment, and, as such, the targets evaluated, and overall assessment approach requires refinement.

For the cell stress panel, the relevance of the models chosen for a given safety decision including proficiency on human-relevant metabolism is key.

There is a consensus that the use of cells, engineered tissues or multi-organ systems of human origin is preferred; however, how to ensure the test systems chosen provide enough biological coverage has not yet been determined. To achieve this, one possibility is to consider a multitude of cells derived from different tissues, whilst another is to focus more on the receptors and pathways expressed in those cells rather than their site of origin. These two perspectives are not mutually exclusive, and it is paramount that the cell types chosen are useful to explore changes at the pathway level. Expectations are to work with cells that have the relevant pathways and features for addressing a variety of modes of action (non-specific and/or specific targets) and express the targets that represent a safety concern. It is also important to consider the biotransformation capacity of the test system used, and whether metabolites of concern are expressed.

Alignment on the extent and relevance of biological coverage useful to inform the safety decisions is needed to build confidence and could be helpful for developing some framework/guidance on the use of such data in the context of NGRA. Transparent documentation of the choice of the tools and reporting of data, and where possible and appropriate standardization of the tools used will contribute to building regulatory confidence.

### **7.3. Being explicit about the level of confidence in the assessment**

The level of confidence in a risk assessment is a consequence of limitations (*e.g.* in assumptions, instrumentation and/or knowledge) and biases and therefore, uncertainty is impossible to eliminate. However, by being cognizant of and by characterizing and documenting the sources and magnitude of limitations and biases, it is possible to take steps to increase confidence and to refine the models used for risk assessment. For example, understanding the reproducibility of individual assays should affect how data are both generated (*e.g.* ensuring a sufficient number of replicates) and interpreted (*e.g.* by applying appropriate statistical analysis to describe the confidence in the data). Furthermore, some of the confidence surrounding the application of NGRA is related to the relative lack of complete NGRA examples, and experience amongst safety assessors of applying NAMs to decision-making. These types of limitations can be overcome with adequate access to examples of NGRA and training materials. Understanding and acknowledging the limitations regarding animal models and traditional approaches to risk assessment are also critical so that in the future, NAMs may improve accuracy and increase confidence by providing a more complete understanding of the response to a xenobiotic.

Defining acceptable levels of confidence prior to data generation in NGRA approaches is rarely straightforward. These approaches should use existing information and, therefore, directly depend on the quality of the databases used for hazard characterization and on the accuracy of exposure estimates. In addition, when filling data gaps, using an iterative approach also entails using various methodologies bearing their specific intrinsic errors, and *a priori* confidence determination may not always be exhaustive. Similarly, determining an acceptable level of confidence subsequent to data generation may also not be ideal, because *a posteriori* confidence determination may introduce bias (*e.g.* external pressure factors tolerating/imposing a reduction in scientific rigour) and this, in itself, further reduces confidence. Therefore, it is imperative to clearly and explicitly document all identifiable limitations and biases (*e.g.* the overall contribution from different assay systems and *in silico* models to uncertainty and propagation of uncertainty; the biological space coverage in terms

of selected cell types and endpoints). An agreed framework for capturing this information would bring consistency and confidence. Finally, it is worth noting that workshop participants were of the opinion that: 1) using robust and relevant NGRA methods and strategies can help minimize the level of uncertainty, 2) when using an iterative approach, appropriate BERs may be adapted as uncertainty is characterized, and 3) when properly addressing uncertainty, bioactivity measures can be suitable for risk-based decision-making. An area of discussion surrounded what an appropriate BER would be. In many cases this will be model dependent (as is the case when the assessment is based on animal test data). In other words, the BER considered to represent a low-risk scenario will depend on the model system used and the biomarkers measured. For this reason, assay developers need to work together with safety assessors to determine relevant benchmark substance exposures to use to characterize the response of these test systems and their relevance for safety decision-making.

#### **7.4. *Agreed standards for using tools and reporting data***

In traditional toxicological risk assessment, the quality of the data obtained and confidence in the overall assessment may be enhanced by basing protocols on regulatory safety studies performed to standardized test methods such as OECD guidelines, using tests performed to Good Laboratory Practice (GLP) standards.

By their nature, many tools and assays used in NGRA are either new or are not yet routinely applied to consumer safety decision-making and not necessarily covered by OECD guidelines or OECD Harmonised Templates. Workshop participants therefore highlighted that, for scientific acceptance agreed standards are needed for using tools, recording methods or protocols, as well as data analysis and reporting.

One possibility for ensuring standardization and acceptance is for new test methods to undergo formal validation to show that they use sound science and meet regulatory needs. After a test has been

validated, its performance (relevance and reliability) is determined to be acceptable for specific proposed uses. The test methods are accepted by international and national testing organizations and regulatory authorities for hazard and risk assessment based upon their history of use and proven utility (OECD, 2005). OECD has also published a guidance document on good *in vitro* method practices (GIVIMP), which are intended to ensure that *in vitro* data consider good scientific, technical, and quality practices (OECD, 2018).

It should be acknowledged that although formal validation may be desired, in some cases other methods of demonstrating relevance and reliability of an assay or approach may be needed. Formal validation is time consuming and may not be appropriate for molecular techniques that do not have a correlate with existing *in vivo* data. The evolution of some novel assays such as micro-physiological systems (MPS), high-throughput screening (HTS) and/or high-content screening (HCS) and HTTr has been rapid, and the specific questions addressed using these techniques may differ greatly from one assessment to another. Furthermore, approaches such as read-across cannot be validated, but best practice should always be followed in terms of the quality of the underpinning data and clear explanation of why the read-across is considered robust. Where it is not possible or appropriate to apply a formal validation process to encourage acceptance and standardization, this should not impede the application of new approaches in an accepted fit-for-purpose context. However, because the information generated from these test methods may be critical for decision-making and can be useful at several levels in the risk assessment process, agreed standards are of the utmost importance. For example, workshop participants agreed that HTTr data play a key role in NGRA, and a lot of the discussion focussed on how to determine which cells, engineered tissues or multi-organ systems to use, how to derive a PoD from these data, and how to report it. Clear scientific and regulatory guidance on these topics was identified as a key immediate need for global acceptance of the overall NGRA approach. Indeed, following the workshop, steps have been made to develop a template for

the description of *in vitro* test data to bring rigour and allow use in regulatory decision-making (Krebs *et al.*, 2019).

Furthermore, such guidance and templates would enable assessors to demonstrate the transparency of the assessment and to allow any decision-making reviewer to understand the data and reasoning behind an assessment, to replicate it, and confirm the same conclusions as those outlined in the original analysis in accordance with the ICCR principles of NGRA (Dent *et al.*, 2018).

#### **7.5. Ability to distinguish adaptation from adversity**

With the types of NGRA discussed at this workshop aiming at being protective rather than being predictive of apical effects, PoDs and defining human protection levels rely on biological activity provided by NAMs. Still, there is a desire to better understand how biological activity relates to apical toxicological activity/outcome in humans. For example, when the BER is large enough to assure that a given safety decision is sufficiently protective, there may be no need to distinguish between adaptation and adversity. In scenarios where the BER is small, further investigation is needed to refine it and where necessary, reduce uncertainty by increasing mechanistic understanding. Distinguishing adaptation from adversity is therefore one of the options available for refining the risk assessment.

Some workshop participants were concerned that the ‘protective not predictive’ approach may be too conservative to be useful for decision-making. In other words, there was concern that low tier *ab initio* assessments based on comparison of a broad set of NAM data with internal exposures calculated using PBK modelling may provide small BERs. This situation would trigger further investigation on the correlation between ‘biological activity’ and ‘toxicological activity/outcome’ to find a way to distinguish between adaptive/reversible responses, for example using quantitative adverse outcome pathways (AOPs). AOPs are specific cases where knowledge is available and permits linking molecular and/or cellular events to adverse outcomes. The AOP framework is still under development and, therefore, it cannot be the sole resource for discriminating adaptation from adversity at this time.



Furthermore, development of new quantitative AOPs for use in this context is extremely challenging, and the cost-benefit of such an undertaking may not often be a viable option for a new ingredient used solely in cosmetics.

A further option is to define potential adversity as signals indicating significant perturbations that would reasonably be expected to lead to an adverse outcome. This will require careful benchmarking of the test system using relevant substance exposures known to cause adverse effects in humans, alongside those that are presumed safe.

Overall, the ability to distinguish between a truly adaptive and adverse effect was considered to represent a potential barrier to the application of NGRA to bioactive substances that are systemically available.

#### **7.6. Updated risk assessment workflow**

By working through a tiered and iterative workflow such as the SEURAT-1 *ab initio* approach (Berggren *et al.*, 2017; see Figure 3) or a similar framework, risk assessors can determine whether new data need to be generated (entirely through new *in vitro* testing and *in vitro*-to-*in vivo* extrapolation) or whether there is already sufficient information to make a decision. Even though the current approach developed with a focal point on the risk assessment of cosmetic ingredients is useful (and it is general enough to be suitable to assess chemicals, exposure scenario and endpoint), there appears to be a need for a workflow designed specifically to handle the data and decisions required for the risk assessment of cosmetic ingredients and to guide assessors through the process.

Workshop participants indicated that this framework should be *non-prescriptive, guided, and context-driven* and should outline the basic components deemed essential to gather the data required for decision-making. This new framework would most likely be inspired by the SEURAT-1 *ab initio* workflow and, to help guide risk assessors, would expand the various steps of this workflow by proposing relevant NAMs at each step of the iterative process, and crucially the decision points that

indicate follow up is necessary. Without being prescriptive, the new NGRA framework would nonetheless need to provide an extensive coverage of the tools and approaches helpful for data mining, data generation, traditional and novel dataset integration, as well as propose an intuitive user interface designed specifically to meet the needs of safety assessors. Some workshop participants proposed the development of a decision tree, whereas others likened their vision of the ideal interface to the popular online tax filing platforms that require users to answer questions, and then provide users with a fit-for-purpose way to walk through the workflow (*e.g.* users input all data gathered about a given substance, and the workflow outputs the required next steps).

Furthermore, problem formulation is critically important when determining which data streams are required for different types of decision-making; accordingly, the decision context (*e.g.* screening, classification, prioritization, hazard characterization, risk characterization, risk assessment) needs to be clearly defined from the start of the assessment. In addition, because this new NGRA framework is intended to be useful for decision-making across different decision contexts, all logical aspects involved in the iterative process would need to be transparently and explicitly documented. This would allow assessors to handle situations involving substances with non-specific toxicity (*e.g.* AOP is unknown/unclear), and to more easily interpret complex, high-content (*e.g.* 'omics) assays.

The helpful tools and approaches used in both systemic case studies presented during the workshop should not be considered as the only ones that would be useful in the risk assessment of cosmetic ingredients, but rather as a subset of tools/approaches that can generally be used. For example, the US EPA's next generation blueprint of computational toxicology (Thomas *et al.*, 2019) proposes to evaluate hazard using, among others, automated, cost-efficient, HTTr (whereby the effects of a given chemical on gene expression are determined using RNA sequencing-based multiplexed readouts) and phenotypic profiling (whereby cultured cells are labelled using multiple fluorescent probes, and high-content imaging is used to quantify changes in subcellular parameters).

Because the use of NGRA is still in its early stages, the risk assessment community has relatively limited experience. Therefore, to obtain a better understanding of how the iterative approach supports decision-making, additional case studies would help in the development, stress-testing and acceptance of the NGRA framework. Finally, our community will have to 1) determine which types of data can be trusted to obtain useful bioactivity readouts; 2) reach a consensus on appropriate data that can be confidently used to implement a new NGRA framework allowing decision-making; 3) benchmark the usefulness of the new NGRA framework against traditional approaches [to encourage acceptance and to expedite adoption/implementation, the new framework needs to be at least as useful as (and may potentially become better and more transparent than) traditional approaches, even if there are knowledge gaps]; 4) recognize that these data will be expected to evolve over time as knowledge advances; and 5) accept that a transition period will be required for safety assessors and regulators to understand and ultimately accept NGRA.

### **7.7. More case studies**

Although not available at the time of the workshop, the authors of both the coumarin and the phenoxyethanol case studies subsequently performed a comparison to see how conservative the NGRA approach presented was compared with the traditional safety assessment. The NGRA for the coumarin case study was found to be at least as protective as the assessment based on traditional (animal-based) approaches (Baltazar *et al.*, 2020), as was the NGRA for phenoxyethanol (OECD, 2021a). This corroborates other analyses that have shown that NAM-derived PoDs are often more conservative than animal-derived PoDs (Paul Friedman *et al.*, 2020). As more complete examples of *ab initio* NGRA case studies are made available, confidence in the application of NAMs for safety decision-making will grow, as will understanding of the limitations of NGRA. Workshop participants were unanimous in their request for access to more systemic NGRAs, including examples where the conclusion would be expected to be high risk. If NGRA fails to protect against substance exposures that we know would be a consumer safety concern more work needs to be done to ensure the NAMs

used are relevant, robust, and provide sufficient coverage. These types of assessments would therefore provide confidence that exposure-led, hypothesis driven NGRA approaches are able to distinguish between high and low risk exposures, or between potent toxicants and inert substances. A critical step in the transition from traditional to NGRA approaches is a comparison of how protective the safety decision made using each approach is, as well as to illustrate when the tools/approaches available for NGRA offer advantages and refinements over traditional testing and assessment approaches.

## 8. CONCLUSIONS

1. The common features of NGRA for systemic effects include a robust assessment of consumer exposure using PBK modelling, including metabolite formation, broad biological coverage using HTTr and specific assays covering interactions of concern. Alignment on a 'base set' of data (which may be modified depending on *e.g. in silico* alerts or existing knowledge) will help to build confidence in NGRA.
2. Seven areas **(7.1-7.7)** were identified to help to make NGRA useful for cosmetic ingredients. This requires investment, especially in further case studies, alongside frameworks to guide performance and reporting of the assessment and its strengths and limitations. These case studies should elaborate scenarios/problem formulations frequently encountered by industry and regulatory safety assessors including those where a 'high risk' conclusion would be expected. This will provide confidence that the tools and approaches can reliably discern differing levels of concern or risk.

## 9. ACKNOWLEDGEMENTS

We dedicate this work to the memory of our friend, colleague and co-author Marcio Lorencini who sadly lost his life to COVID-19 during the finalization of this report. The case study presenters (PLC, HK and GO) would like to thank their respective case study teams for performing the work that made these discussions possible.

## 10. REFERENCES

- Ackley, D., Abernathy, M., Chaves, A., Damiano, B., Delaunois, A., Foley, M., ... Vargas, H. (2019). Current nonclinical in vivo safety pharmacology testing enables safe entry to first-in-human clinical trials: The IQ consortium nonclinical to clinical translational database. *Journal of Pharmacological and Toxicological Methods*, 99, 106595. <https://doi.org/10.1016/j.vascn.2019.05.152>
- Allen, T. E. H., Goodman, J., Gutsell, S., & Russell, P. J. (2014). Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. *Chemical Research in Toxicology*, 27(12), 2100–2112. <https://doi.org/10.1021/tx500345j>
- Amaral, R., Da Silva, Ansell, J., Boisleve, F., Dent, M., Hatao, M., ... Weiss, C. (2017). REPORT FOR INTERNATIONAL COOPERATION ON COSMETICS REGULATION Regulators and Industry Joint Working Group (JWG): Integrated Strategies for Safety Assessment of Cosmetic Ingredients: Part 1. Retrieved August 13, 2021 from [https://www.iccr-cosmetics.org/downloads/topics/iccr-11\\_jwg\\_integrated\\_strategies\\_integrated\\_strategies\\_for\\_safety\\_assessments\\_of\\_cosmetic\\_ingredients\\_-\\_part\\_i.pdf](https://www.iccr-cosmetics.org/downloads/topics/iccr-11_jwg_integrated_strategies_integrated_strategies_for_safety_assessments_of_cosmetic_ingredients_-_part_i.pdf)
- Amaral, R., Ansell, J., Boisleve, F., Cubberley, R., Dent, M., Hatao, M., ... Weiss, C. (2018). REPORT FOR INTERNATIONAL COOPERATION ON COSMETICS REGULATION Regulators and Industry Joint Working Group (JWG): Integrated Strategies for Safety Assessment of Cosmetic Ingredients: Part 2. Retrieved September 3, 2019, from <https://iccr->

[cosmetics.org/files/8315/4322/3079/ICCR Integrated Strategies for Safety Assessment of Cosmetic Ingredients Part 2.pdf](https://cosmetics.org/files/8315/4322/3079/ICCR_Integrated_Strategies_for_Safety_Assessment_of_Cosmetic_Ingredients_Part_2.pdf)

- Andersen, M. E., Clewell, H. J., Carmichael, P. L., & Boekelheide, K. (2011). Can case study approaches speed implementation of the NRC report: “toxicity testing in the 21st century: a vision and a strategy?”. *ALTEX*, 28(3), 175–182. <https://doi.org/10.14573/altex.2011.3.175>
- Baltazar, M. T., Cable, S., Carmichael, P. L., Cubberley, R., Cull, T. A., Delagrange, M., ... Westmoreland, C. (2020). A next generation risk assessment case study for coumarin in cosmetic products. *Toxicological Sciences*, 176(1), 236–252. <https://doi.org/10.1093/toxsci/kfaa048>
- Berggren, E., White, A., Ouedraogo, G., Paini, A., Richarz, A.-N., Bois, F. Y., ... Mahony, C. (2017). Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods. *Computational Toxicology*, 4, 31–44. <https://doi.org/10.1016/J.COMTOX.2017.10.001>
- Blackburn, K. L., Carr, G., Rose, J. L., & Selman, B. G. (2020). An interim internal Threshold of Toxicologic Concern (iTTC) for chemicals in consumer products, with support from an automated assessment of ToxCast™ dose response data. *Regulatory Toxicology and Pharmacology*, 114. <https://doi.org/10.1016/j.yrtph.2020.104656>
- Bowes, J., Brown, A. J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., & Whitebread, S. (2012). Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nature Reviews Drug Discovery*, 11(12), 909–922. <https://doi.org/10.1038/nrd3845>
- Browne, P., Judson, R. S., Casey, W. M., Kleinstreuer, N. C., & Thomas, R. S. (2015). Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model. *Environmental Science and Technology*, 49(14), 8804–8814. <https://doi.org/10.1021/acs.est.5b02641>
- Catlett, N. L., Bargnesi, A. J., Ungerer, S., Seagaran, T., Ladd, W., Elliston, K. O., & Pratt, D. (2013). Reverse causal reasoning: Applying qualitative causal knowledge to the interpretation of high-

- throughput data. *BMC Bioinformatics*, 14(1), 340. <https://doi.org/10.1186/1471-2105-14-340>
- Clewell, H. J., Thomas, R. S., Kenyon, E. M., Hughes, M. F., Adair, B. M., Gentry, P. R., & Yager, J. W. (2011). Concentration- and time-dependent genomic changes in the mouse urinary bladder following exposure to arsenate in drinking water for up to 12 weeks. *Toxicological Sciences*, 123(2), 421–432. <https://doi.org/10.1093/toxsci/kfr199>
- De Abrew, K. N., Kainkaryam, R. M., Shan, Y. K., Overmann, G. J., Settivari, R. S., Wang, X., ... Daston, G. P. (2016). Grouping 34 chemicals based on mode of action using connectivity mapping. *Toxicological Sciences*, 151(2), 447–461. <https://doi.org/10.1093/toxsci/kfw058>
- De Abrew, K. N., Shan, Y. K., Wang, X., Krailler, J. M., Kainkaryam, R. M., Lester, C. C., ... Daston, G. P. (2019). Use of connectivity mapping to support read across: A deeper dive using data from 186 chemicals, 19 cell lines and 2 case studies. *Toxicology*, 423, 84–94. <https://doi.org/10.1016/j.tox.2019.05.008>
- Dent, M. P., Amaral, R. T., Da Silva, P. A., Ansell, J., Boisleve, F., Hatao, M., ... Kojima, H. (2018). Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. *Computational Toxicology*, 7, 20–26. <https://doi.org/10.1016/J.COMTOX.2018.06.001>
- Desprez, B., Dent, M., Keller, D., Klaric, M., Ouédraogo, G., Cubberley, R., ... Mahony, C. (2018). A strategy for systemic toxicity assessment based on non-animal approaches: The Cosmetics Europe Long Range Science Strategy programme. *Toxicology in Vitro*, 50, 137–146. <https://doi.org/10.1016/j.tiv.2018.02.017>
- Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W., & Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, 95(1), 5–12. <https://doi.org/10.1093/toxsci/kfl103>
- Ellison, C. A., Blackburn, K. L., Carmichael, P. L., Clewell, H. J., Cronin, M. T. D., Desprez, B., ... Worth,

- A. (2019). Challenges in working towards an internal threshold of toxicological concern (iTTC) for use in the safety assessment of cosmetics: Discussions from the Cosmetics Europe iTTC Working Group workshop. *Regulatory Toxicology and Pharmacology*, *103*, 63–72.  
<https://doi.org/10.1016/j.yrtph.2019.01.016>
- Ellison, C. A., & Wu, S. (2020). Application of structural and functional pharmacokinetic analogs for physiologically based pharmacokinetic model development and evaluation. *Regulatory Toxicology and Pharmacology*, *114*, 104667. <https://doi.org/10.1016/j.yrtph.2020.104667>
- Farmahin, R., Gannon, A. M., Gagné, R., Rowan-Carroll, A., Kuo, B., Williams, A., ... Yauk, C. L. (2019). Hepatic transcriptional dose-response analysis of male and female Fischer rats exposed to hexabromocyclododecane. *Food and Chemical Toxicology*, *133*.  
<https://doi.org/10.1016/j.fct.2018.12.032>
- Farmahin, R., Williams, A., Kuo, B., Chepelev, N. L., Thomas, R. S., Barton-Maclaren, T. S., ... Yauk, C. L. (2017). Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. *Archives of Toxicology*, *91*(5), 2045–2065.  
<https://doi.org/10.1007/s00204-016-1886-5>
- Gilbert, P., Beveridge, E. G., & Crone, P. B. (1977a). Effect of phenoxyethanol on the permeability of *Escherichia coli* NCTC 5933 to inorganic ions. *Microbios*, *19*(75), 17–26.
- Gilbert, P., Beveridge, E. G., & Crone, P. B. (1977b). The lethal action of 2-phenoxyethanol and its analogues upon *Escherichia coli* NCTC 5933. *Microbios*, *19*(76), 125–141.
- Gwinn, W. M., Auerbach, S. S., Parham, F., Stout, M. D., Waidyanatha, S., Mutlu, E., ... DeVito, M. J. (2020). Evaluation of 5-day In Vivo Rat Liver and Kidney With High-throughput Transcriptomics for Estimating Benchmark Doses of Apical Outcomes. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, *176*(2), 343–354. <https://doi.org/10.1093/toxsci/kfaa081>
- Hatherell, S., Baltazar, M. T., Reynolds, J., Carmichael, P. L., Dent, M., Li, H., ... Middleton, A. M. (2020). Identifying and Characterizing Stress Pathways of Concern for Consumer Safety in Next-



Generation Risk Assessment. *Toxicological Sciences*, 176(1), 11–33.

<https://doi.org/10.1093/toxsci/kfaa054>

Houck, K. A., Dix, D. J., Judson, R. S., Kavlock, R. J., Yang, J., & Berg, E. L. (2009). Profiling Bioactivity of the ToxCast Chemical Library Using BioMAP Primary Human Cell Systems. *Journal of Biomolecular Screening*, 14(9), 1054–1066. <https://doi.org/10.1177/1087057109345525>

Krebs, A., Waldmann, T., Wilks, M. F., Van Vugt-Lussenburg, B. M. A., Van der Burg, B., Terron, A., ... Leist, M. (2019). Template for the description of cell-based toxicological test methods to allow evaluation and regulatory use of the data. *ALTEX*, 36(4), 682–699.

<https://doi.org/10.14573/altex.1909271>

Labib, S., Williams, A., Guo, C. H., Leingartner, K., Arlt, V. M., Schmeiser, H. H., ... Halappanavar, S. (2016). Comparative transcriptomic analyses to scrutinize the assumption that genotoxic PAHs exert effects via a common mode of action. *Archives of Toxicology*, 90(10), 2461–2480.

<https://doi.org/10.1007/s00204-015-1595-5>

Labib, Sarah, Williams, A., Yauk, C. L., Nikota, J. K., Wallin, H., Vogel, U., & Halappanavar, S. (2016). Nano-risk Science: Application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. *Particle and Fibre Toxicology*, 13(1).

<https://doi.org/10.1186/s12989-016-0125-9>

Lewis, R. W., Billington, R., Debryune, E., Gamer, A., Lang, B., & Carpanini, F. (2002). Recognition of Adverse and Nonadverse Effects in Toxicity Studies. *Toxicologic Pathology*, 30(1), 66–74.

<https://doi.org/10.1080/01926230252824725>

Lobenhofer, E., Cui, X., Bennett, L., Cable, P., Merrick, B., Churchill, G., & Afshari, C. (2004).

Exploration of low-dose estrogen effects: Identification of No Observed Transcriptional Effect Level (NOTEL). *Toxicologic Pathology*, 32(4), 482–492.

<https://doi.org/10.1080/01926230490483324>

Lynch, J. J., Van Vleet, T. R., Mittelstadt, S. W., & Blomme, E. A. G. (2017). Potential functional and

pathological side effects related to off-target pharmacological activity. *Journal of Pharmacological and Toxicological Methods*, 87, 108–126.

<https://doi.org/10.1016/j.vascn.2017.02.020>

Macfarlane, M., Jones, P., Goebel, C., Dufour, E., Rowland, J., Araki, D., ... Scheel, J. (2009). A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: Skin irritation. *Regulatory Toxicology and Pharmacology*, 54(2), 188–196.

<https://doi.org/10.1016/j.yrtph.2009.04.003>

MHLW. (2011). *Guidebook on the Manufacture and Sales of Cosmetics and Quasi-Drugs* (pp. 159-182 [In Japanese]). pp. 159-182 [In Japanese]. Tokyo: Yakuji Nippo Ltd.

Monticello, T. M., Jones, T. W., Dambach, D. M., Potter, D. M., Bolt, M. W., Liu, M., ... Kadambi, V. J. (2017). Current nonclinical testing paradigm enables safe entry to First-In-Human clinical trials: The IQ consortium nonclinical to clinical translational database. *Toxicology and Applied Pharmacology*, 334, 100–109. <https://doi.org/10.1016/j.taap.2017.09.006>

Moxon, T. E., Li, H., Lee, M. Y., Piechota, P., Nicol, B., Pickles, J., ... Baltazar, M. T. (2020). Application of physiologically based kinetic (PBK) modelling in the next generation risk assessment of dermally applied consumer products. *Toxicology in Vitro*, 63, 104746.

<https://doi.org/10.1016/j.tiv.2019.104746>

OECD. (2021a). *Case study on the use of Integrated Approaches for Testing and Assessment for the systemic toxicity of phenoxyethanol when included at 1% in a body lotion*. Series on Testing and Assessment, No 349, Environment, Health and Safety, Environment Directorate, OECD.

[https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/CBC/MONO\(2021\)35&docLanguage=En](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/CBC/MONO(2021)35&docLanguage=En)

OECD. (2021b). *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.

<https://www.oecd.org/chemicalsafety/risk-assessment/guidance-document-on-the-characterisation-validation-and-reporting-of-physiologically-based-kinetic-models-for-regulatory-purposes.pdf>

Paul Friedman, K., Gagne, M., Loo, L.-H., Karamertzanis, P., Netzeva, T., Sobanski, T., ... Thomas, R. S. (2020). Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization. *Toxicological Sciences*, 173(1), 202–225.

<https://doi.org/10.1093/toxsci/kfz201>

Qutob, S. S., Chauhan, V., Kuo, B., Williams, A., Yauk, C. L., McNamee, J. P., & Gollapudi, B. (2018). The application of transcriptional benchmark dose modeling for deriving thresholds of effects associated with solar-simulated ultraviolet radiation exposure. *Environmental and Molecular Mutagenesis*, 59(6), 502–515. <https://doi.org/10.1002/em.22196>

Rogiers, V., Benfenati, E., Bernauer, U., Bodin, L., Carmichael, P., Chaudhry, Q., ... Worth, A. (2020, April 30). The way forward for assessing the human health safety of cosmetics in the EU - Workshop proceedings. *Toxicology*, Vol. 436, p. 152421.

<https://doi.org/10.1016/j.tox.2020.152421>

Simmons, S. O., Fan, C. Y., & Ramabhadran, R. (2009). Cellular stress response pathway system as a sentinel ensemble in toxicological screening. *Toxicological Sciences*, Vol. 111, pp. 202–225.

<https://doi.org/10.1093/toxsci/kfp140>

Thomas, R. S., Allen, B. C., Nong, A., Yang, L., Bermudez, E., Clewell, H. J., & Andersen, M. E. (2007). A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure. *Toxicological Sciences*, 98(1), 240–248.

<https://doi.org/10.1093/toxsci/kfm092>

Thomas, R. S., Bahadori, T., Buckley, T. J., Cowden, J., Deisenroth, C., Dionisio, K. L., ... Williams, A. J. (2019). The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency. *Toxicological Sciences*, 169(2), 317–332.

<https://doi.org/10.1093/toxsci/kfz058>

Thomas, R. S., Clewell, H. J., Allen, B. C., Wesselkamper, S. C., Wang, N. C. Y., Lambert, J. C., ...

Andersen, M. E. (2011). Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicological Sciences*, *120*(1), 194–205.

<https://doi.org/10.1093/toxsci/kfq355>

Troutman, J. A., Rick, D. L., Stuard, S. B., Fisher, J., & Bartels, M. J. (2015). Development of a physiologically-based pharmacokinetic model of 2-phenoxyethanol and its metabolite phenoxyacetic acid in rats and humans to address toxicokinetic uncertainty in risk assessment. *Regulatory Toxicology and Pharmacology*, *73*(2), 530–543.

<https://doi.org/10.1016/j.yrtph.2015.07.012>

US EPA. (2014). *Next Generation Risk Assessment: Incorporation Of Recent Advances In Molecular, Computational, And Systems Biology (Final Report)*, EPA/600/R-14/004. Washington DC: US Environmental Protection Agency.

[https://ofmpub.epa.gov/eims/eimscomm.getfile?p\\_download\\_id=520491](https://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=520491)

Wambaugh, J. F., Hughes, M. F., Ring, C. L., MacMillan, D. K., Ford, J., Fennell, T. R., ... Thomas, R. S. (2018). Evaluating in vitro-in vivo extrapolation of toxicokinetics. *Toxicological Sciences*, *163*(1), 152–169. <https://doi.org/10.1093/toxsci/kfy020>

Whitebread, S., Dumotier, B., Armstrong, D., Fekete, A., Chen, S., Hartmann, A., ... Urban, L. (2016, August 1). Secondary pharmacology: screening and interpretation of off-target activities – focus on translation. *Drug Discovery Today*, Vol. 21, pp. 1232–1242.

<https://doi.org/10.1016/j.drudis.2016.04.021>

Wu, S., Fisher, J., Naciff, J., Laufersweiler, M., Lester, C., Daston, G., & Blackburn, K. (2013).

Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chemical Research in Toxicology*, *26*(12), 1840–1861. <https://doi.org/10.1021/tx400226u>