



September 19, 2019

Animal-Free Safety Assessment (AFSA) Scoping Document: Resources supporting implementation of non-animal methods in consumer product safety assessment

Prepared for:

Catherine Willett, PhD
Senior Director, Science and
Regulatory Affairs
Research & Toxicology Department
Humane Society International
kwillett@hsi.org
(240) 599-6785

Submitted by:

Rebecca Clewell, PhD
Principal Consultant
21st Century Tox Consulting, LLC.
rclewell@21stCenturyTox.com
919-551-8100

Introduction

Background

Humane Society International (HSI) is partnering with industry groups and other interested parties to facilitate the acceptance and adoption of non-animal methods in chemical safety assessments. With a goal of ending animal testing of cosmetics and cosmetic ingredients by 2023, HSI is working to develop training and education packages for various stakeholders including regulators, members of industry, academic researchers and informed consumers, including consulting and contracting service providers. The focus of this training will be on how to make robust consumer safety decisions with target audiences anticipated to be regulators, industry safety assessors (including SMEs), safety and regulatory compliance consultants, academic researchers and students, and informed consumers. Significant advances have been made over the last decade in the development of tools and technologies for non-animal safety assessments. With this current AFSA collaboration, the ultimate goal is to facilitate incorporation of these non-animal methods into real-world decision making by improving access to knowledge, education and testing resources.

Scope of Work

To ensure that the resulting training materials are comprehensive, targeted to the needs of the audience, and include current state of the science, this document aims to determine the current scientific tools and approaches to support non-animal safety assessments, as well as the ability of stakeholders to access these tools and approaches, with a particular emphasis on those that may be useful for cosmetic ingredients. Such tools and approaches include decision frameworks (e.g. Risk21, etc.), tools for evaluating and predicting chemical kinetics *in vivo* (PBPK) and translating *in vitro* results to *in vivo* exposures (IVIVE), *in vitro* testing methods, and *in silico* predictive tools for bioactivity and read-across. In evaluating access to tools and technologies, we have attempted to identify vendors (contract research organizations, etc.) that can provide non-animal testing services to industry and computational platforms that are available either as freeware or through licensing agreements. We will also identify academic toxicological networks including both brick and mortar and online toxicology programs that include New Approach Methods (NAMs) and/or Next Generation Risk Assessment (NGRA) and those that do not. The initial focus of this work was on services and products that are available in the United States and the European Union. Scoping work was performed via literature search (PubMed) and web search, as well as through contact with various industry, government and academic stakeholders via vendor directories for societies such as the Society of Toxicology and personal contacts.

General Methods

Web Searches

Google and GoogleScholar were used for web-based searching. Search terms were selected based on the topic of interest and results were sorted by relevance. For each search term, each entry from the first 5 pages of results were evaluated for relevancy to the topic. 5 pages of results were selected based on preliminary searches for educational resources that were examined. In this preliminary effort, relevance of results decreased substantially after the first 3 pages of results. For this reason, a consistent cut-off of 5 pages (~50 results) was set for

subsequent searches. In most cases, the first two search term combinations yielded > 80% of the hits that we determined to be relevant to the subject of interest.

Examples of search strings for *in vitro* testing: “*in vitro* test cosmetics”, “*in vitro* assay cosmetic toxicity”, “*in vitro* testing skin irritation”, *etc.*

Examples for ADME modeling: “intrinsic clearance QSAR”, “parent chemical clearance prediction”, “parent chemical clearance *in silico*”, “ADME model platforms”, *etc.*

PubMed was used to search for publications on approaches to NGRA or safety assessment strategies using non-animal methods with consumer products/cosmetics. Results from these searches are included in the report and attached worksheet.

Identifying Resources by Expert Opinion.

In addition to the objective web-based searches, we also specifically searched for organizations that were known to us based on our experience. We further solicited additional information from colleagues and known experts in the field. These results were added to the worksheet when they were relevant. However, it is important to note that technologies found in PubMed, but not provided as services or products by vendors (*i.e.*, academic research exercises) are not included in this document. Only those technologies that are readily available for use by all stakeholders are included here.

Results

Safety Assessment Paradigms

RISK21

The International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI) created the Risk Assessment in the 21st Century (RISK21) project in 2009 with goal of improving human health assessment. The project is a multi-sector, international initiative that has continued to evolve over the last 10 years. The RISK21 framework (Figure 1) is a problem formulation-based approach that emphasizes the need to incorporate consideration of exposure at all stages of a risk assessment. Ultimately, evaluation of risk is based on a comparison of the dose where toxicological response (Figure 1A) occurs and the likely exposure (Figure 1B) to the population of interest. This ratio of expected exposure to toxicological dose is also known as the margin of exposure (MoE) and is an important concept in traditional risk assessments, as well as in several of the frameworks that have been developed for risk assessments based on non-animal methods. In the RISK21 project, an interactive web-based tool has been developed to provide graphical representation of this MoE (*i.e.*, matrix; Figure 1C).

Importantly, the RISK21 framework is designed to be performed in an iterative and tiered manner, beginning with rapid toxicity and exposure evaluation methods and moving to more data rich, time-intensive methods only as needed to improve accuracy of the predicted MoE. In such a tiered approach, the assumption is that while the rapid methods for estimating toxicological doses and population exposures are more uncertain than more in-depth methods, if the MoE is sufficiently large, it may not be necessary to spend added time or money on

refining the estimates. However, for cases where the MoE is low, more in depth testing may help to refine the MoE or allow for better clarity on appropriate conditions of use. Thus, tiered approaches help to prioritize allocation of resources to those chemicals that are more likely to have public health consequences, and also allows for rapid decisions to be made for chemicals that are unlikely to pose unacceptable risk to the public. While the RISK21 framework, as published, includes a tier for collection of *in vivo* data, the framework is generalizable and can be applied to a non-animal approach by replacing *in vivo* tests with targeted, organotypic *in vitro* testing. The RISK21 framework has been described in detail in publications (Embry et al., 2014) and the utility of the framework for decision-making has been demonstrated for a number of prototype chemicals and with a number of exposure scenarios. A full list of citations can be found on the RISK21 website (<https://hesiglobal.org/risk-assessment-in-the-21st-century-risk21/?kp2=2#publications2>).

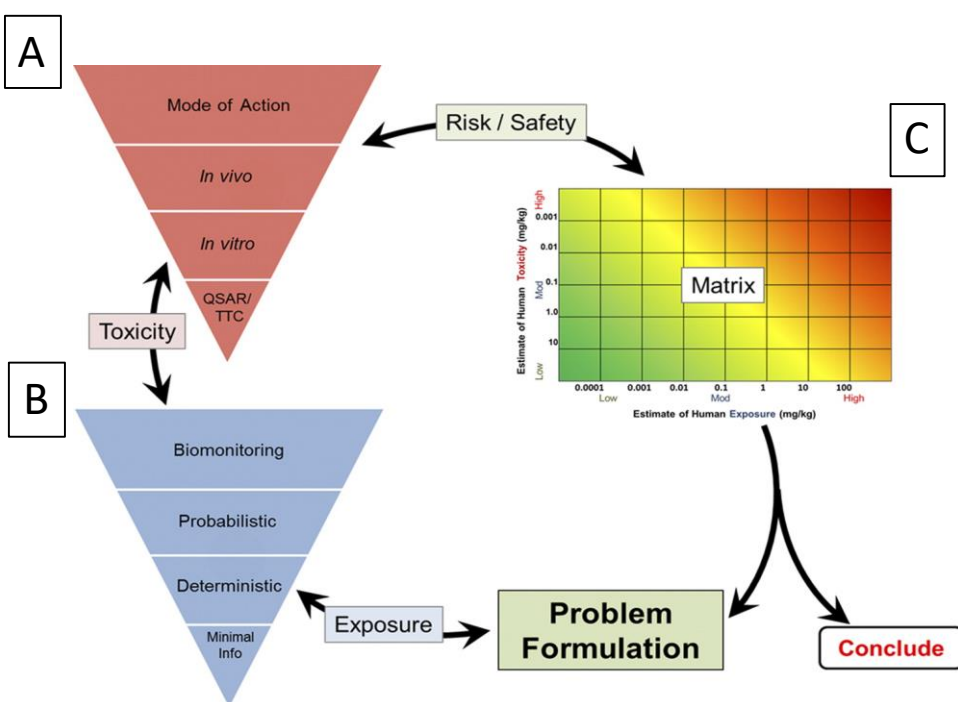


Figure 1. RISK21 framework. (Reproduced from Embry et al., 2014).

Margin of Exposure-based Prioritization

In 2016, the United States congress passed the Frank R. Lautenberg Chemical Safety for the 21st Century Act (LCSA), amending the long-standing Toxic Substances Control Act (TSCA) to require more comprehensive and efficient evaluation of the safety of new and existing chemicals by the EPA. This amendment included language for the use of scientifically valid non-animal test methods for the first time in the United States' legislation. A key driver of this bill was the existence of thousands of chemicals in commercial use that have not been systematically tested for human or environmental toxicity, as a result of the grand-father clause in the original TSCA. In the United States, oral or inhalation reference dose assessments only exist for 366 chemicals, approximately 0.5% of the TSCA substance inventory of over 67,000 chemical agents (USEPA 2016; 2018). It is clear that traditional animal-based testing cannot

possibly meet the need for information on potential toxicity of existing chemicals, let alone the ever-increasing numbers of new chemistries being developed.

In the absence of exposure and toxicity data for most commercial compounds, the U.S. EPA's National Center for Computational Toxicology (NCCT) has been developing computational tools for the rapid estimate of exposure and chemical bioactivity, and the Tox21 consortium has tested bioactivity of thousands of chemicals in hundreds of *in vitro* assays (<https://tox21.gov>). To date, over 4,500 chemicals have been tested in 500-700 ToxCast assays (https://www.epa.gov/sites/production/files/2019-01/documents/toxcast_factsheet_dec2018.pdf), and more than 6,000 chemicals have been tested in approximately 70 Tox21 assays (<https://tox21.gov>).

To help support the push for more rapid chemical prioritization and testing under the LCSA, USEPA's NCCT is promoting the use of these high throughput exposure and bioactivity methods for prioritization of chemical testing based on predicted MoE (Figure 2). *In vitro* measured potencies can be converted to predicted *in vivo* points of departure using a computational technique known as *in vitro* to *in vivo* extrapolation (IVIVE) (Rotroff et al., 2010). IVIVE is a rapid physiologically based pharmacokinetic (PBPK) modeling approach that uses *in vitro* or *in silico* derived estimates of hepatic clearance, serum protein binding and oral absorption to predict the oral equivalent dose (OED) associated with steady state blood concentrations equal to the media concentration associated with *in vitro* bioactivity for a chemical. Thus, *in vitro* measured bioactivity coupled with IVIVE can be used together with rapid exposure estimates to derive MoEs without time and resource intensive animal studies (Rotroff et al., 2010; Wetmore et al., 2012; 2015). While there are significant uncertainties in the estimated MoEs using these rapid estimation methods, the ranking of compounds by MoE enables prioritization of chemicals based on the likelihood that human exposure could lead to toxicity. The focus here is not on defining chemical mode of action or the nature of potential chemical-induced toxicity, but on using the power of multiple data streams to predict exposure levels at which chemicals are likely to be bioactive.

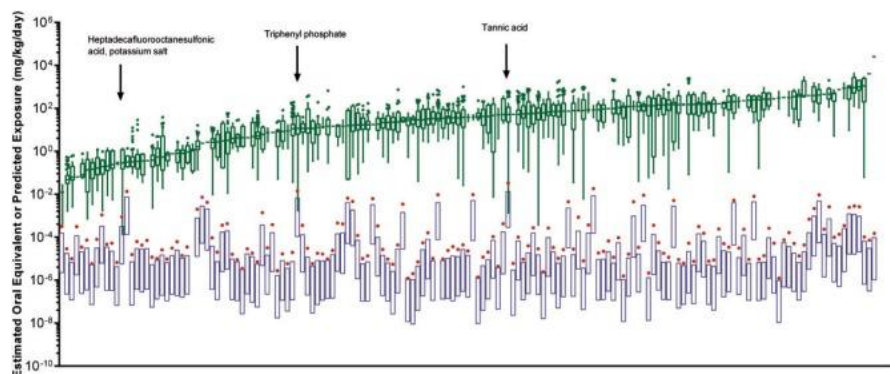


Figure 2. Comparison of predicted margins of exposure for 163 ToxCast Phase II chemicals. Distributions of the OEDs across approximately 700 *in vitro* assays for each chemical are depicted as box-and-whisker plots, presented with exposure predictions derived from (Wambaugh et al., 2014). Data are ordered from lowest to highest median OEDs. A full list of chemicals and supporting data are provided in Supplementary Table S4. Predicted exposures are represented by floating bars, with the lower bar value representing the geometric mean and the upper bar the upper 95% confidence limit around the mean. The red filled circle denotes the upper 95% confidence limit derived for the most highly exposed (MHE)

population for that chemical. Arrows indicate chemicals with AERs <1. Figure reproduced from Wetmore et al., 2015.

Recently, an international governmental collaboration (Accelerating the Pace of Chemical Risk Assessment; APCRA) was formed with the aim of bringing together governmental entities engaged in risk assessment to discuss progress and barriers in applying new tools to prioritization, screening, and quantitative risk assessment of differing levels of complexity and to pursue increased collaboration with the goal of accelerating the pace of chemical risk assessment. Toward this end, APCRA is performing a number of collaborative case studies. A retrospective case study compares points of departure derived from traditional animal studies with those that would be derived from *in vitro* assays (ToxCast screening) (https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NCCT&dirEntryId=345649). Such efforts provide practical tests for the utility of these *in vitro* based MoE approaches for supporting risk-based decisions, which will be required for increasing acceptance of NAMs for regulatory purposes.

Adverse Outcome Pathways (AOPs), Integrated Approaches to Testing and Assessment (IATA) and Defined Approaches (DAs)

The process of developing an AOP begins with identification of a toxicological outcome for which sufficient mode of action data exists to define a series of causal events that can lead to a particular phenotype. The resulting AOP depicts a linear set of events beginning with a molecular initiating event (ligand-receptor binding, damage to a cellular component, etc.), followed by subsequent cellular, tissue and organism level events (key events) that can lead to a particular toxicological (adverse) outcome. The included events must be causal, not merely associated. However, initiation of one or more of the key events may not result in the organism-level effect. Thus, an AOP defines necessity, but not sufficiency, of the key events in the progression to an adverse outcome. Importantly, development of such a schematic inherently defines the key biological processes that should be considered in a testing approach.

IATA are defined as pragmatic, science-based testing approaches for chemical hazard assessment that rely on integrated analysis of existing information with generation of new data. IATA are typically iterative and designed to answer a defined question in a specific regulatory context, e.g. “Is chemical A capable of inducing hazard B in species C?”. The overall assessment process within IATA is based on weight-of-evidence (WoE), which relies heavily on expert judgment in assigning the weight of the various pieces of information. IATA can range from informal grouping and read-across type activities to more structured, prescriptive approaches, such as those recently adopted for eye and skin irritation and skin sensitization.

The development of a skin sensitization IATA exemplifies the use of an AOP approach to define appropriate *in vitro* assays for chemical assessment (Patlewicz et al., 2014). Through expert opinion and literature review, a series of events required for skin sensitization were defined - from the molecular initiating event (covalent modification of proteins/haptenization) to activation of oxidative stress response and release of inflammatory factors in dendritic cells and keratinocytes activation of T cell response, which ultimately leads to a hypersensitivity of the skin to allergen exposure (OECD, 2012). Based on this AOP, *in silico* QSAR and read-across methods and several *in vitro* assays that recapitulated key events (Nrf2 activation,

haptization, etc.) were recommended as non-animal alternatives to *in vivo* assays (Figure 3). Currently, the recommended strategy for testing of chemicals (OECD GD 256) is almost entirely animal free, with the exception of activation of the systemic immune response, for which there is no replacement for the murine local lymph node assay (OECD TG 429, 442A/B). A similar approach was used in the development of the OECD published IATA for skin and eye irritation (OECD GD 203, 263). And by the USEPA to develop a testing strategy for estrogen disruptors – defining a pathway and leveraging available assays to evaluate hazard (Browne et al., 2015). The IATA Case Studies Project was launched in 2015 under the revised Cooperative Chemicals Assessment Programme (CoCAP) to increase experience with the use of IATA by developing case studies as examples of testing strategies for regulatory use. Currently, there are 15 case studies published or under review (<http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm#Project>).

Defined Approaches (DAs) are testing regimens that are rule-based and are not influenced by expert judgment (Casati et al., 2018). The fixed nature of DAs is meant to facilitate their consideration under the OECD mutual acceptance of data, whereas IATAs are designed to be flexible and adaptable to particular regional requirements or regulatory statutes. These defined approaches provide rigorous, well-documented and highly curated support for specific decisions that must be made for various chemical classifications. However, these DAs are not generalizable to scenarios outside the defined context. In this case, the IATA describes skin sensitization as the toxicological outcome of interest, and the DAs define specific tests that must be performed to support decisions about the sensitizing potential of a compound, which is only one component of a risk assessment for cosmetics. Under the broader skin sensitization IATA, there are currently 12 DAs in which data generated by non-animal methods are evaluated by means of a fixed data interpretation procedure. Together, AOPs, IATA, and DAs provide a link from the science of *in vivo* toxicology to *in silico* and *in vitro* test methods, ultimately paving the way to predicting hazard based on non-animal methods.

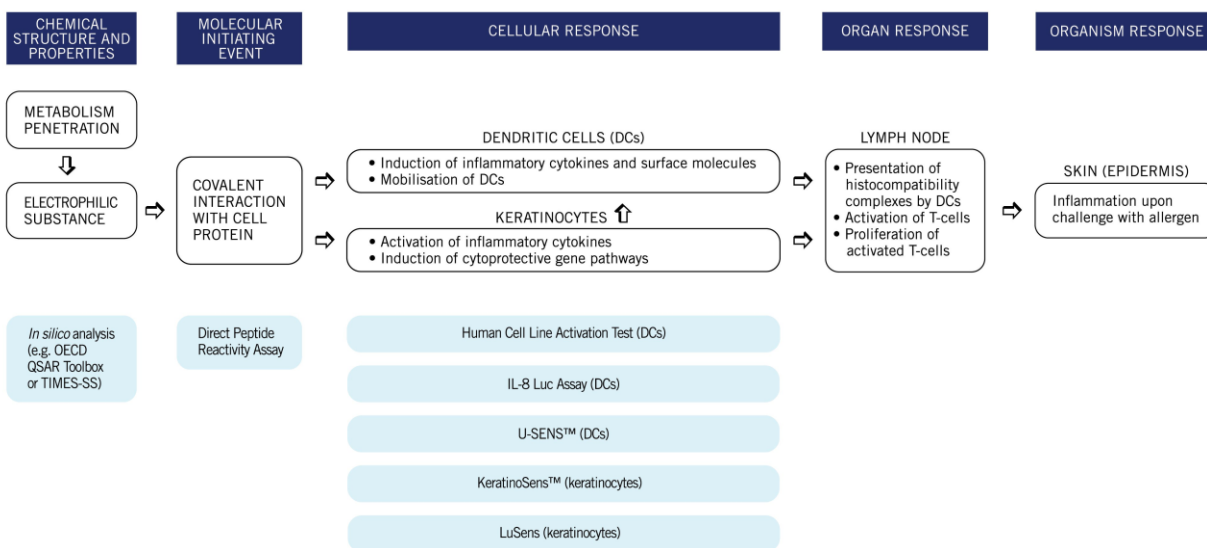


Figure 3. AOP for skin sensitization and *in silico/in vitro* tests corresponding to key events. This figure is reproduced from Peta International Science Consortium Ltd (<https://www.piscld.org.uk/skin-sensitisation/>).

Cosmetics Europe Approach to Non-animal Testing in Cosmetics

In their 2017 report on non-animal approaches to safety assessment of cosmetic products, Cosmetics Europe defined 5 key components of risk assessment for cosmetics:

- (1) eye irritation and severe eye damage (ocular toxicity),
- (2) genotoxicity/mutagenicity,
- (3) skin sensitization (includes skin irritation),
- (4) absorption, distribution, metabolism, and elimination (ADME), and
- (5) systemic toxicity (including carcinogenicity, reproductive toxicity, and repeat dose toxicity).

(For report, see: https://www.cosmeticseurope.eu/files/1215/0245/3923/Non-animal_approaches_to_safety_assessment_of_cosmetic_products.pdf).

Of these, *in vitro* test methods are particularly well-advanced for ocular toxicity, genotoxicity and skin sensitization (Figure 4). *In vitro* and *in silico* methods are also rapidly improving for predicting *in vivo* ADME. However, a strategy for predicting systemic toxicity is still very much in development. Systemic toxicity includes considerations such as multiple potential target organs and potential for repeat dose toxicity. Currently large-scale research efforts are underway to develop a strategy for dealing with systemic toxicity, including the transcriptomic approaches at USEPA NCCT, the chemi-informatics approaches being explored by Cosmetics Europe, and the large-scale assay development in EU-ToxRisk. However, the lack of validated assays for systemic toxicity currently precludes defining an *in vitro* based testing strategy.

Several efforts using case studies, including the large scale, multi-year, EU integrated project to develop *in vitro* testing strategies for acute toxicity (<http://www.AcuteTox.org>) have suggested that cytotoxicity (the neutral red uptake; NRU) is currently the most predictive assay for acute toxicity (LD50) (Sjöström et al., 2008). However, a repeat dose, truly systemic *in vitro* approach will require substantial development. In the meantime, approaches such as read-across and Thresholds of Toxicological Concern (TTC) may be used to estimate safe levels of exposure based on chemical structure. In fact, in recent efforts, the ICCVAM acute toxicity workgroup (ATWG) organized a global collaboration to build predictive *in silico* models for acute oral systemic toxicity. Resulting computational models were found to yield consensus predictions which demonstrated excellent performance when compared to the animal data, indicating that this approach to acute toxicity may soon provide a reasonable replacement for animal tests (Kleinstrauber et al., 2018).

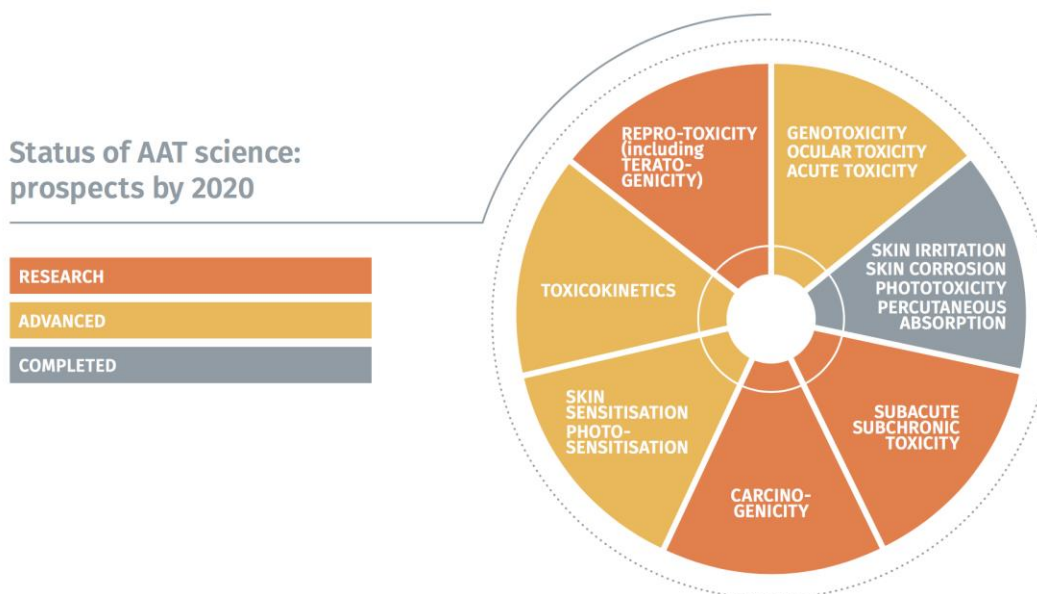


Figure 4. Status of development of nonanimal test methods for key components of a risk assessment for personal care products. Reproduced from Cosmetics Europe report on Non-animal Approaches to Safety Assessment of Cosmetic Products (https://www.cosmeticseurope.eu/files/1215/0245/3923/Non-animal_approaches_to_safety_assessment_of_cosmetic_products.pdf).

SEURAT Conceptual Framework for Non-animal Safety Assessment

SEURAT-1 was an EU funded public-private research consortium from 2011-2016 aimed at addressing the long-term strategic target of "Safety Evaluation Ultimately Replacing Animal Testing (SEURAT)". The overall project combined the research efforts of over 70 European universities, public research institutes and companies in an effort to develop knowledge and technology building blocks required for the development of solutions for the replacement of current repeated dose systemic toxicity testing *in vivo* used for the assessment of human safety.

From this effort, a conceptual framework for a rational integrated assessment strategy emerged from designing the case studies and is discussed in the context of international developments focusing on alternative approaches for evaluating chemicals using the new 21st century tools for toxicity testing. This framework is shown below in Figure 5.

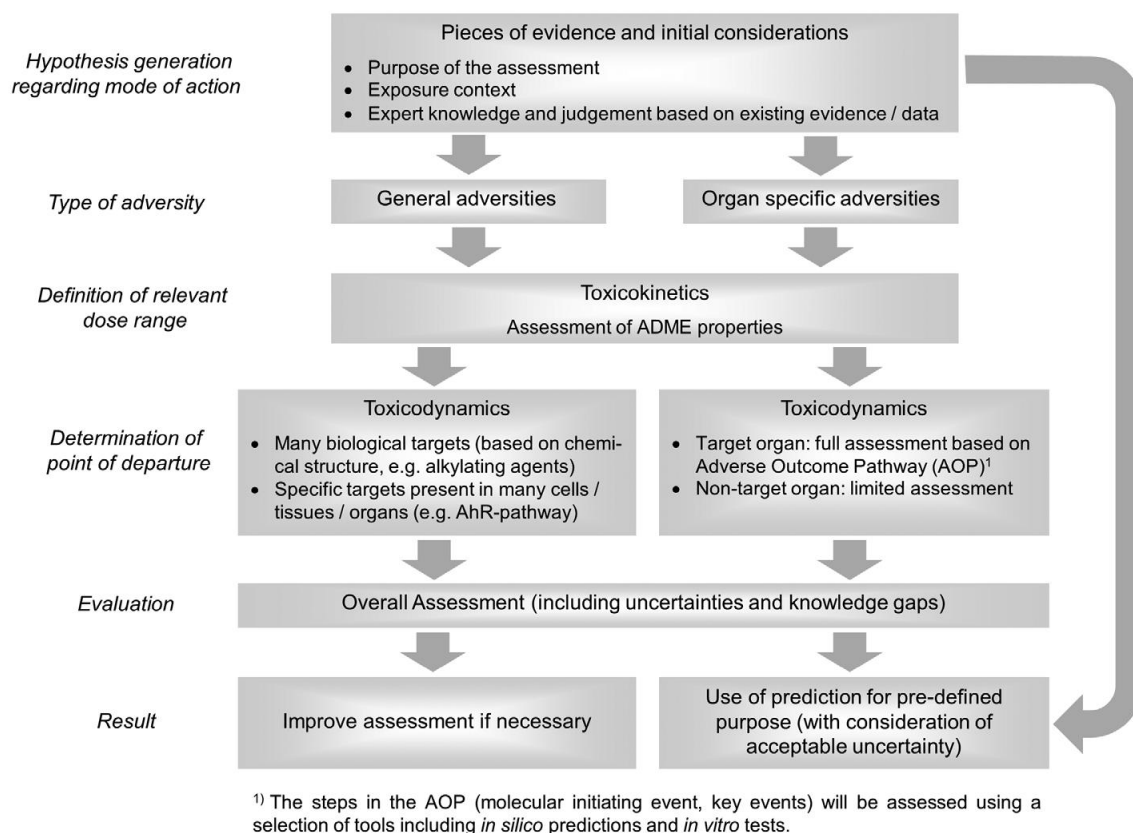


Figure 5. SEURAT conceptual framework for animal-free safety assessment.

This framework highlights the importance of early consideration of exposure, as well as hypothesis driven research. This framework also highlights the importance of defining, early on, the degree of confidence needed for the prediction. The evidence that is incorporated into the assessment may then be chosen based on the degree of certainty that is required for the specific decision context (Gocht et al., 2015).

International Cooperation on Cosmetics Regulation (ICCR) Principles Underlying the Next Generation Risk Assessment Approach (NGRA)

The International Cooperation on Cosmetics Regulation (ICCR), a working group of scientists representing both regulatory and industry stakeholders, developed a series of principles for incorporating NAMs into a Next Generation Risk Assessment (NGRA). In this context, an NGRA is defined as an exposure-led, hypothesis driven risk assessment approach that integrates *in silico*, *in chemico* and *in vitro* data streams. The ICCR concluded that each NGRA will need to be customized to the available data, the intended use and likely exposure. Thus, a prescriptive list of tests to assure the safety of a chemical in cosmetics does not exist and would not be appropriate. Instead, the ICCR developed nine principles that should guide NGRAs in general.

- 1) The overall goal of an 'NGRA' is a human safety assessment, and therefore must be human relevant.
- 2) The assessment is exposure-led.
- 3) The assessment should be hypothesis-driven.
- 4) The assessment is designed to prevent harm.
- 5) The assessment must follow an appropriate appraisal of all existing information.
- 6) The assessment should be conducted by using a tiered and iterative approach.
- 7) Robust and relevant methodologies and testing strategies must be used to ensure confidence in the validity of the safety assessment.
- 8) Sources of uncertainty should be well-characterized and discussed.
- 9) All data, assumptions, methodologies and software used should be clearly documented and available for independent review.

These principles are described in depth in a publication by Dent and coauthors (2018) as well as the full ICCR report (https://www.iccr-cosmetics.org/files/4715/0824/0761/ICCR-11_JWG_Integrated_Strategies_Integrated_Strategies_for_Safety_Assessments_of_Cosmetic_Ingredients_-_Part_I.pdf).

Comparing the Identified Risk Assessment Frameworks

The current review effort focused on frameworks that have been proposed for incorporation of NAMs into risk assessment. While a number of different terminologies are used to describe these efforts (e.g., 21st century risk assessment, next generation risk assessment, non-animal risk assessment), the unifying characteristic is the move away from complete reliance upon animal testing toward the incorporation of – or even total dependence on – *in vitro* and *in silico* methodologies. Another characteristic that is reiterated across these frameworks, is the use of tiered evaluations. For the most part, these frameworks avoid prescriptive testing regimens and instead focus on the guiding principles that can ensure robust, tractable, health-protective assessments. These frameworks have built in flexibility because, as a scientific community, we are still in the process of developing many of the test methods that will be needed to fully transition to non-animal methods. As more sophisticated *in silico* and *in vitro* models are developed, we must be able to quickly incorporate these methodologies in risk assessment practice. Further, with the ability to tailor *in vitro* testing to the problem at hand, it is no longer expedient to take a “one size fits all” approach to testing of chemicals with vastly different properties and uses. Thus, a risk assessment framework that makes the best use of the available technologies must have built-in flexibility.

The Risk21 and MoE-based prioritization approaches focus heavily on early consideration of exposure and exposure (or risk) based decision making with the goal of ensuring human relevance of resulting safety decisions. Early consideration of exposure ensures that resources are used most efficiently by focusing on chemicals that are likely to be associated with exposures at bioactive concentrations. Risk-based decisions steer away from chemical categorization and focus instead on quantitative evaluations of chemical bioactivity that can predict safe levels of human exposure, regardless of the type of toxicity.

The Cosmetics Europe roadmap, as well as the IATA and associated DAs, contrast with these MoE-based approaches in that they are more focused on hazard identification. Cosmetics

Europe attempts to define the toxicological endpoints that must be addressed in order to have sufficient confidence in a risk assessment for personal care products. AOPs and IATAs are even more hazard driven – they define a series of biological events that must occur in order for a chemical to cause a particular toxicological event. DAs translate the information from an IATA to testing protocols. DAs, as opposed to the high level, flexible risk assessment frameworks such as RISK21 and ICCR’s NGRA, are highly prescriptive testing approaches designed to answer specific risk assessment questions in specific contexts.

The ICCR-NGRA is a comprehensive approach that focuses on exposure-led assessments, quantitative methods, incorporates tiered testing strategies and focuses on the use of biologically relevant nonanimal methods. Here, we use this approach as a guidance for the collection of *in silico* and *in vitro* resources that can be used to support a NGRA for personal care products.

Availability of NAMs to Support Non-animal Risk Assessments

In order to be useful for supporting a chemical risk assessment for consumer products, a test method must be highly vetted, with a well-defined domain of applicability. There is much debate surrounding the future of test validation, as it will be impossible for traditional validation and test guideline development practices to keep pace with the rapid development of new technologies (Burgdorf et al., 2019; Judson et al., 2013; Whelan and Eskes, 2016). However, for the purposes of the current effort, we focused on tests that currently have been validated by ECVAM, have published OECD test guidelines (TGs), guidance documents (GDs), or are modifications of tests with published OECD test guidelines. Exceptions are the GARD and SENS-IS assays. These assays were included in the results, as they have been approved to undergo evaluation by ECVAM (<https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/eurl-ecvam-status-report-development-validation-and-regulatory-acceptance-alternative-3>). For those categories identified below, however, there were at times additional non-guideline assays that are either widely used by the research community (and deemed likely candidates for future guideline development), or were vital to implementing modeling approaches (e.g., parent chemical clearance for PBPK/IVIVE). Providers of these non-guideline assays were recorded in the supplementary spreadsheet when found during the literature search, but were not a focus of the primary research effort.

Models such as those used for biokinetics/ADME, read-across and TTC do not undergo the same validation and guideline publication process as *in vitro* assays. Thus, we include in this effort any models that are offered as services by CROs or consulting groups, or those that are provided to the public by government agencies, such as the USEPA’s HHTK package, or the OECD Toolbox.

Based on the recommendations of the Cosmetics Europe report, together with the ICCR NGRA focus on exposure-led assessments, we identified the following *in vitro* and *in silico* methods to be of particular interest for the current work:

1. Ocular toxicity:
 - Reconstructed human cornea-like epithelium (RhCE) *in vitro* test method (e.g. EpiOcular™, SkinEthic™) – OECD TG 492
 - Bovine Corneal Opacity and Permeability (BCOP) test method – OECD TG 437
 - Short time exposure (STE) *in vitro* method – OECD TG 491

- Isolated Chicken Eye (ICE) *in vitro* test method (using otherwise discarded materials) – OECD TG 438
- Fluorescein leakage (FL) *in vitro* test method – OECD TG 460
- 2. Skin toxicity:
 - Skin irritation/corrosion:
 - *In vitro* membrane barrier test, Corrositex – OECD TG 435
 - *In vitro* skin corrosion, reconstructed human epidermis (RHE) test – OECD TG 431
 - *In vitro* skin irritation, reconstructed human epidermis (RHE) test – OECD TG 439
 - Skin sensitization:
 - ARE-Nrf2 luciferase test method (KeratinoSens™ and LuSens) – OECD TG 442D
 - Direct peptide reactivity assay (DPRA) – OECD TG 442C
 - Human Cell Line Activation Test (h-CLAT) – OECD TG 442E
 - IL-8 Luc assay – OECD TG 442E
 - U937 Skin Sensitization Test (U-SENS™) – OECD TG 442E
 - GARD
 - SENS-IS
 - Phototoxicity
 - Cytotoxicity with photoinduction – OECD 432
- 3. Genotoxicity/carcinogenicity:
 - AMES bacterial reverse mutation test – OECD TG 471
 - *In vitro* micronucleus test – OECD TG 487
 - *In vitro* mammalian chromosome aberration test – OECD TG 473
 - *In vitro* mammalian cell gene mutation test (HPRT) – OECD TG 476
 - *In vitro* mammalian cell gene mutation test using thymidine kinase gene, aka mouse lymphoma assay – OECD TG 490
 - Unscheduled DNA synthesis in mammalian cells *in vitro* – OECD 482
 - *In vitro* cell transformation assays – OECD GD 214, 231
- 4. ADME (Biokinetics)
 - Skin penetration
 - *In vitro* diffusion method – OECD TG 428
 - *In vitro* metabolism
 - Parent chemical clearance using microsomes, S9, hepatocytes or recombinant enzymes
 - ADME prediction
 - User friendly platforms* that predict PK parameters based on structure
 - PBPK modeling
 - User-friendly modeling platforms*
 - IVIVE
 - User-friendly modeling platforms*
- 5. Systemic toxicity
 - Acute toxicity

- 3T3 neutral red uptake cytotoxicity (or similar cytotoxicity assays)– OECD TG 129
 - Threshold of toxicological concern (TTC)
 - User-friendly modeling platforms*
 - Read-across
 - User-friendly modeling platforms*
6. Exposure
- Predictive models of human exposure (focusing on personal care products)

**"User friendly" refers to modeling/computational tools that do not require users to have specialized expertise in scripting languages or de novo model development.*

In vitro Testing

Laboratories and consulting companies that provide previously identified NAMs as services were identified via web search as described in the General Methods and through personal contact with experts in the field. Tests that are minor variations on the OECD guideline tests listed above are also included. A detailed list of providers and tests is provided in the attached Excel worksheet (see attached worksheet: *In vitro* tests). A summary of the findings and brief highlights of recent important developments in assay methodologies is provided below.

Ocular Toxicity

In 2017, the OECD released a Guidance Document (GD 263) describing an IATA for eye irritation and serious eye damage, which recommends a tiered approach to classifying potential chemicals based on existing data, QSAR and read-across, and *in vitro* testing (OECD, 2017). A number of *in vitro* assays exist for ocular toxicity. Of these, most focus on cytotoxicity as the endpoint of interest, as it is the primary, mechanistic role in determining the chemical-induced eye damage (OECD, 2017). The models range from traditional monolayer cultures, to chicken eggs, to chicken, bovine or porcine eyes obtained as a by-product from abattoirs (otherwise discarded tissues), and 3-D human reconstructed tissue (reconstructed human corneal epithelium; RhCE). These *in vitro* models are described in detail elsewhere (Wilson et al., 2015).

There are OECD Test Guidelines for 5 assays considered in the IATA:

- Reconstructed human cornea-like epithelium (RhCE) *in vitro* test method – OECD TG 492
- Bovine Corneal Opacity and Permeability (BCOP) test method – OECD TG 437
- Short time exposure (STE) *in vitro* method – OECD TG 491
- Isolated Chicken Eye (ICE) *in vitro* test method – OECD TG 438
- Fluorescein leakage (FL) *in vitro* test method – OECD TG 460

Test Guideline 492 is based on commercial three-dimensional RhCE tissue constructs that are produced using either primary human epidermal keratinocytes (*i.e.*, EpiOcular™; MatTek Corp.), human immortalized corneal epithelial cells (*i.e.*, SkinEthic™ HCE/S; Episkin), or primary human corneal epithelial cells (*i.e.*, LabCyte CORNEA-MODEL24 or MCTT HCETM). As these RhCE tissue constructs are similar to the *in vivo* corneal epithelium three-dimensional structure and are produced using cells from the species of interest (human), they represent a significant step

forward in fully replacing animal use (including animal tissues) for chemical toxicity testing. 3D reconstructed tissue models for the skin and lung are also highly advanced and are being incorporated into safety assessment paradigms. These models are discussed in later sections of this document.

A summary of the findings for providers of these ocular toxicity assays are shown in Table 1. Details are provided in the attached worksheet. In addition to the above guideline assays, data were also recorded for widely used assays including the Irritaction® assay (*In vitro* International), the porcine corneal opacity and permeability (PCOP) assay, and the chorioallantoic membrane vascular assay (CAMVA, Het-CAM; performed in chicken eggs).

Table 1. Providers of *in vitro* assays for ocular irritation and corrosion.

OECD Test Guideline	Assay	Number of Providers
492	Eye irritation; Reconstructed human cornea-like epithelium (RhCE)	13
437	Bovine corneal opacity and permeability (BCOP)	1
491	Short time exposure (STE)	4
438	Isolated chicken eye (ICE)	3
460	Fluorescein leakage (FL)	1
	Other: Irritaction, PCOP, CAMVA, Het-CAM	7

Skin Irritation and Sensitization

Similar to ocular toxicity, skin irritation (reversible damage) and corrosion (serious damage) are primarily driven by acute cytotoxicity. Thus, the IATA for skin irritation and corrosion is similar to that of ocular toxicity (OECD, 2017B), and the OECD Guidance Document recommends a tiered approach to classifying potential chemicals based on existing data, QSAR and read-across, and *in vitro* testing (OECD, 2017B).

Three assays exist with OECD test guidelines for skin irritation and corrosion:

- *In vitro* membrane barrier test, Corrositex – OECD TG 435
- *In vitro* skin corrosion, reconstructed human epidermis (RhE) test – OECD TG 431
- *In vitro* skin irritation, reconstructed human epidermis (RhE) test – OECD TG 439

Several reconstructed human epidermis (RhE) models exist and three are included in the OECD Test Guidelines 431 and 432. The RhE test system uses human derived non-transformed keratinocytes as the cell source to reconstruct an epidermal model with organotypic cytostructure, including a base membrane, proliferating keratinocytes and a stratum corneum with an intact barrier function. These models exist at the air:liquid interface, and therefore allow realistic exposures, as well as research on barrier function and metabolism. The four models included in TG 431 and 439 are:

- EpiSkin (L'Oréal, France)
- EpiDerm (MatTek, US)
- SkinEthic (L'Oréal, France)
- LabCyte EPI-Model (J-TEC, Japan).

For the purposes of this report, we pooled the data for all RhE models in the attached spreadsheet.

Six assays exist with OECD test guidelines for skin sensitization, and two assays are scheduled for validation by ECVAM:

- ARE-Nrf2 luciferase test method (KeratinoSens and LuSens) – OECD TG 442D
- Direct peptide reactivity assay (DPRA) – OECD TG 442C
- Human Cell Line Activation Test (h-CLAT) – OECD TG 442E
- IL-8 Luc assay – OECD TG 442E
- U937 Skin Sensitization Test (U-SENS) – OECD TG 442E
- GARD
- SENS-IS

Finally, a guideline assay exists for evaluating the potential of chemical agents to induce toxicity upon subsequent exposure to light (TG 432), by comparing the cytotoxicity of a test compound when tested with and without additional exposure to a non-toxic dose of UVA and visible light. This assay may use standard *in vitro* test cell lines (such as mouse NIH 3T3 fibroblasts), or more fit for purpose human skin models, including skin-derived cell lines, primary skin cell monocultures or even 3-D RhE models. Some variation also exists in how cytotoxicity is measured (neutral red uptake, MTT, etc). For this review, we combined results of the different models/readouts and present all suppliers of phototoxicity assays as a single entry in the attached spreadsheet.

Guideline assays for other endpoints (DNA damage, skin sensitization, etc) have also been adapted to evaluate synergistic effects between chemical and light exposure and are used toward predicting overall risk of a compound. These modifications were not explicitly included in the search conducted for this review. However, if such a service was advertised by one of the providers that was evaluated, a note was made in the attached worksheet.

A summary of the findings for providers of these skin toxicity assays are shown in Table 2. Details are provided in the attached worksheet.

Table 2. Providers of *in vitro* assays for skin irritation and sensitization.

OECD Test Guideline	Assay	Number of Providers
	Irritation	
435	Corrositex	7
431	<i>In vitro</i> skin corrosion, RHE	15
439	<i>In vitro</i> skin irritation, RHE	17
	Sensitization	
442D	KeratinoSens, LuSens	10
442C	Direct peptide reactivity assay (DPRA)	11
442E	h-CLAT	7
442E	IL-8 Luc assay	0
442E	U-SENS	1
	GARD	2
	SENS-IS	4
	Phototoxicity	

432	Phototoxicity (monoculture, RhE)	15
-----	----------------------------------	----

Genotoxicity and Carcinogenicity

While tests have been used for more than four decades to screen for potential carcinogens, with *in vitro* methods dating back to the Ames bacterial mutation assay in the 1970's (Ames et al., 1972), there is as of yet no prescribed replacement for *in vivo* cancer assays. Nonetheless, several well-validated assays exist that are used in early screening for carcinogenic potential. Genotoxicity assays measure the potential for chemical-induced DNA damage. Other assays test the ability of chemicals to induce mutation (permanent change to genetic code), or unscheduled DNA synthesis (indicative of uncontrolled proliferation), or cellular transformation (transformation from normal to neoplastic cells), which are key characteristics of carcinogenicity.

Eight assays exist with OECD Test Guidelines or Guidance Documents:

- AMES bacterial reverse mutation test – OECD TG 471
- *In vitro* micronucleus test – OECD TG 487
- *In vitro* mammalian chromosome aberration test – OECD TG 473
- *In vitro* mammalian cell gene mutation test (HPRT) – OECD TG 476
- *In vitro* mammalian cell gene mutation test using thymidine kinase gene, aka mouse lymphoma assay – OECD TG 490
- Unscheduled DNA synthesis in mammalian cells *in vitro* – OECD 482
- *In vitro* cell transformation assays – OECD GD 214, 231

Of the above assays, the micronucleus assay (TG 487) and chromosome aberration test (TG 473) measure genotoxicity (DNA damage). The other assays measure mutation, unscheduled DNA synthesis and *in vitro* cell transformation as indicated in their names – these were considered to be carcinogenesis assays for purposes of this report.

A summary of the findings for providers of these assays are shown in Table 3. Details are provided in the attached worksheet. In addition to the above guideline assays, data were collected for widely used assays including the *in vitro* comet assay, the p-H2AX (DNA repair center) assay, the ToxTracker assay, which measures DNA damage along with other mechanistic information.

Recently, the micronucleus assay has been adapted for use in the EpiDerm™ 3D reconstructed human skin model (Roy et al., 2016). This reconstructed human skin micronucleus (RSMN) assay provides a human relevant, organotypic model for evaluation of DNA damage. Providers for this assay are also included in the attached worksheet if they were observed while performing the primary search for guideline assays.

Table 3. *In vitro* assays for genotoxicity and carcinogenicity.

OECD Test Guideline	Assay	Number of Providers
471	Ames bacterial mutation assay	18
487	<i>In vitro</i> micronucleus	15

473	<i>In vitro</i> chromosome aberration	11
476	<i>In vitro</i> gene mutation test (HPRT)	6
490	<i>In vitro</i> mammalian cell gene mutation assay	10
482	Unscheduled DNA synthesis in mammalian cells	2
Guidance document 241, 231	<i>In vitro</i> cell transformation	0
	Other: pH2AX, comet, etc.	11

ADME

The only OECD Guideline assay related to ADME is the skin penetration/*in vitro* diffusion method (OECD TG 428), which provides information on the absorption of a chemical into and across the skin. This information is vital for understanding both site of contact (skin) dosimetry and systemic absorption (bioavailability). These parameters are important input for PBPK models of dermal exposures (Clewell & Clewell, 2008).

While no OECD Test Guideline currently exists for metabolism, parent chemical clearance is a key parameter for PBPK and IVIVE models (Clewell & Clewell, 2008; Rotroff et al., 2010; Wetmore et al., 2012; 2015; Yoon et al., 2015). While QSAR models exist, they are currently not well parameterized for nonpharmaceutical compounds (Moreau et al., In Preparation). As such, the majority of IVIVE and PBPK efforts use *in vitro* metabolism studies in hepatocytes, S9 fractions, or microsomes to evaluate parent chemical clearance, which then provides the intrinsic clearance parameter needed for ADME models (Rotroff et al., 2010). Because this parameter is key to understanding dosimetry and relating the concentrations of all *in vitro* endpoints to *in vivo* external dose, we included parent chemical clearance (using hepatocytes, S9 or microsomes) in our search of providers.

A summary of the findings for providers of these assays are shown in Table 4. Details are provided in the attached worksheet.

Table 4. *In vitro* assays for ADME.

OECD Test Guideline	Assay	Number of Providers
428	Skin penetration	6
	Parent chemical clearance (hepatocyte, liver S9, liver microsomes)	6

Eventually, IVIVE efforts will need to also account for bioactivation via metabolism and Phase II enzyme depletion, though most current efforts do not include this information. Although not a focus of the current research, we did make note of providers that include these advanced metabolism assays when they were encountered. This additional information is provided in the attached worksheet but is not included in the summary here in the main report.

Systemic Toxicity

As described previously (see section on Cosmetics Europe Approach to Non-animal Testing in Cosmetics), there currently is no comprehensive strategy for predicting systemic toxicity using *in vitro* models, though this is certainly a subject of much research.

Several efforts using case studies, including the large scale, multi-year, EU integrated project to develop *in vitro* testing strategies for acute toxicity (www.AcuteTox.org) have suggested that cytotoxicity (the neutral red uptake; NRU) is currently the most predictive assay for acute toxicity (LD50) (Sjöström et al., 2008). This assay does have an associated OECD Test Guideline (TG 129). We included this Guideline assay, as well as similar assays with different methods of testing cytotoxicity (e.g., MTT or LDH vs. NRU) or different cell lines. A summary of the findings for providers of these assays are shown in Table 5. Details are provided in the attached worksheet.

Table 5. *In vitro* assays for systemic toxicity.

OECD Test Guideline	Assay	Number of Providers
129	Cytotoxicity	16

Ultimately, systemic *in vitro* approach, and particularly one that will consider repeat dose toxicity, will require substantial development. In the meantime, computational approaches such as read-across and TTC may be used to estimate potential toxicological outcomes and safe levels of exposure based on chemical structure. Resources for such computational approaches are discussed in the following section.

**A Note on Lung Toxicity*

While inhalation exposure, and site of contact toxicity (lung irritation, sensitization) are important factors in cosmetics risk assessment, there are currently no *in vitro* models with OECD guidelines for lung toxicity. It should be noted, however, that recent developments in reconstructed airway models provide *in vitro* platforms for testing inhalation toxicity in biologically relevant, human systems. Importantly, these models exist for long periods of time (weeks) at the air:liquid interface, and allow for physiologically relevant exposure scenarios.

These reconstructed models range from single cell type 3-D models to models with multiple cell types, and represent different regions of the airway from the small airway to the muciliary epithelium. These systems use multiple cell types, including club cells, basal cells, goblet cells, ciliated epithelial cells, and fibroblasts, which helps to mimic more phenotypic changes, such as fibrosis, cell death, loss of cilia, etc., that may ensue following chemical exposure. In addition to companies producing complex 3D systems, private researchers are producing bioprinted alveolar models to quad-culture systems and lung-on-a-chip models that include mechanical simulation of the expansion and contraction that occurs in the breathing lung (Benam et al., 2017). These organotypic models provide a path forward in the replacement of animals for inhalation exposure and lung toxicity.

Currently, reconstructed human airway models that are being used for cosmetics, and are provided as services by contract research labs (e.g., Institute for *In vitro* Sciences; IIVS, Charles River Laboratories) are the EpiAlveolar and EpiAirway models from MatTek, and MucilAir and SmallAir from Epithelix. Exposure delivery systems for these models include liquid and aerosolized particles/droplets (using the VitroCell exposure system; <https://www.vitrocell.com>).

In silico Modeling Platforms

QSAR and Read-across for Toxicological Endpoints (Hazard Assessment)

An excellent primer on QSAR and read-across methods (known collectively as “non-test methods”) is available at alttox.org (<http://alttox.org/mapp/emerging-technologies/non-test-approaches-qsars-read-across/>). A structure-activity relationship (SAR) is a qualitative association between a chemical substructure and the chemical’s potential to elicit a specific biological effect. A quantitative structure-activity relationship (QSAR) is a statistical correlation of quantitative parameter derived from chemical structures to a measure of biological activity. (Q)SARs may be based on statistical inference, expert judgment, or some combination of statistics and expert judgement. The benefits of using (Q)SAR approaches include low cost, high throughput, and reduced animal testing. While non-testing approaches are increasingly being used to support chemical safety decisions, the complexity of the models makes evaluation and validation difficult. The European Chemicals Agency (ECHA) has developed guidelines to help facilitate the submission of QSAR models with REACH packages (ECHA, 2016B).

Read-across is a SAR technique for predicting endpoint information for a target substance, by using data from the same endpoint from similar substances. This technique is used in risk assessment for compounds for which there is insufficient toxicity data. Read-across may follow an analogue or category approach. In an analogue approach, the prediction for the target substance is based on available data for a very limited number of structurally-similar substances. In a category approach, the prediction for the target substance is based on available data on several structurally similar substances that are grouped together. The European Chemicals Agency (ECHA) has developed guidelines for the use of read-across approaches in REACH submissions. Generally, only substances that are highly similar in structure may be used for interpolation, but not extrapolation, of toxicological outcomes (ECHA, 2016a; https://echa.europa.eu/documents/10162/13628/raaf_en.pdf).

In an EU LIFE funded effort, the ANTARES (Alternative Non Testing methods Assessed for REACH Substances) project was formed with the goals of reducing the gaps in knowledge concerning the non-testing methods (NTMs, promoting the NTMs for use under REACH, and improving communication about these methods between scientists, regulators and industry. Main activities of this project, which took place from 2009 -2013 were to identify available QSAR models for read-across, assess and validate existing NTMs (in particular (Q)SAR), in order to allow their application for regulatory purposes. More than 250 software models covering 38 REACH endpoints were identified by the ANTARES project. Of these models, 70 were free. The models were classified by prediction endpoints into the following groups: Physico-chemical properties, toxicological (human toxicity), and ecological (ecotoxicity). Toxicological endpoints identified are summarized in Figure 6.

ENDPOINTS	MODELS
8.1. Skin irritation or skin corrosion	
8.1.1. In-vivo skin irritation	+
8.2. Eye irritation	
8.2.1. In-vivo eye irritation	+
8.3. Skin sensitisation	++
8.4. Mutagenicity	
8.4.1. In-vitro gene mutation study in bacteria	++
8.4.2. In-vitro cytogenecity or micronucleus study in mamm. cells	-
8.4.3. In-vitro gene mutation study in mammalian cells	-
8.4.4. In-vivo mutagenicity study	+
8.5. Acute toxicity	
8.5.1. Acute toxicity - by oral route	++
8.5.2. Acute toxicity - by inhalation	+
8.5.3. Acute toxicity - by dermal route	-
8.6. Repeated dose toxicity	+
8.6.1. Short-term repeated dose toxicity (28 d)	+
8.6.2. Sub-chronic toxicity study (90 days)	+
8.6.3. A long-term repeated dose toxicity (> 12 m)	-
8.7. Reproductive toxicity	++
8.7.1. Screening for reproductive-/developmental toxicity	-
8.7.2. Pre-natal developmental toxicity study	-
8.7.3. Two-generation reproductive toxicity study	-
8.8. Toxicokinetics	++
8.8.1. Assessment of the toxicokinetic	-
8.9.1. Carcinogenicity study	++

Figure 6. QSAR models for toxicological endpoints compiled by ANTARES project. “-“ denotes an absence of models, “+” indicates that a few models exist, “++” indicates several models exist.

(Reproduced from ANTARES final report.

http://www.anteslife.eu/files/antes_layman_report_eng_update.pdf).

The list compiled by ANTARES served as a starting point for the current review effort. Additional models were identified via web search. In the intervening years, several important collaborative efforts have focused on improving QSAR models for toxicological endpoints. Such efforts include the development of the OPERA, CERAPP and CoMPARA models for prediction of physico-chemical properties and environmental endpoints, estrogen receptor binding, and androgen receptor binding, respectively (Manganelli et al., 2019; Mansouri et al., 2016; 2018). Other important developments in QSAR and read-across technologies that are more pertinent to the current inquiry include the development and release of the publicly available modeling suites OECD Toolbox, as well as the commercially available REACHAcross™ (<https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm>; <https://www.ulreachacross.com>).

In developing a list of computational tools for the current scoping exercises, we narrowed the search to QSAR and read-across models that predicted our previously determined endpoints of interest: Ocular irritation/corrosion, skin irritation/corrosion/sensitization, genotoxicity/mutagenicity/carcinogenicity, and systemic toxicity including acute and repeat dose lethality. Only user-friendly models that were available for use by the public (either as freeware,

as commercial licenses, or as commercial services) were included in the list. The results can be found in the attached Excel worksheet (see attached worksheet: QSAR Read-across Toxicity).

Threshold of Toxicological Concern (TTC)

The basis of the TTC concept is that “safe levels of exposure” for humans can be determined from chemical structures, even when directly applicable toxicity data are unavailable (Kroes et al., 2005). The method is based on comparing the test chemical structure to the structures in a database of chemicals with experimentally no effect levels (NOELs). A decision tree approach identifies structural alerts and assigns chemicals to one of 3 classes: low (Class I), intermediate (Class II), and high (Class III), which correspond TTCs of 1.5, 9, 30 µg/kg-BW/d, respectively. The TTC values represent the lowest 5th percentile of the distribution of NOAELs for the compounds in the class with *in vivo* data, divided by a safety factor of 100. Extensions of the TTC concept include structural alerts and defined TTC values for specific toxicological outcomes, including genotoxicity, acetylcholinesterase inhibition, reproductive toxicity, etc (Kroes et al., 2005). TTCs are broadly used by industry and regulatory bodies (e.g., U.S. FDA and European Commission) for compounds that are not currently regulated or that lack sufficient data for making risk assessment decisions. These TTC values are based on NOELs from oral dosing data in animals.

ToxTree is a freeware modeling platform that supports TTC estimation for unknown compounds based on chemical structure. The software, which was commissioned by JRC Computational Toxicology and Modelling and developed by Ideaconsult Ltd., allows TTC estimation for single compounds or batches of compounds, using Kroes decision tree, Cramer rules, Cramer rules with extensions, or the new Revised Cramer Decision Tree. The software and associated instruction manuals can be downloaded at <http://toxtree.sourceforge.net/>. This tool, which is easy to use, even for those who are not computational chemists, allows rapid estimation of safe levels of exposure for most chemical structures. The OECD Toolbox also supports TTC estimation using Cramer classification (<http://www.oecd.org/env/ehs/risk-assessment/41833798.pdf>) is freely available.

This method has been specifically evaluated for utility with cosmetics and is used broadly for predictive purposes (Api et al., 2015; Kroes et al., 2007; Yang et al., 2017). The TTC approach has been extended to address dermal (Api et al., 2016; Williams et al., 2016) and inhalation exposures (Schüürmann et al., 2016), and to identify potential skin sensitizers (Safford, 2008), which are of particular use in cosmetics. However, we did not find these extensions in a publicly available modeling platform. There are also ongoing efforts to develop an internal TTC approach, which would be representative of internal exposures (i.e., plasma concentration). An internal TTC (iTTC) would provide threshold values that could be utilized in exposure-based safety assessments. Cosmetics Europe is currently sponsoring research in this area (Ellison et al., 2019).

QSAR Models for ADME

We searched for ADME prediction models, focusing on models that predict parent chemical clearance. Currently, many of the available models have been built for, and vetted against, pharmaceutical compounds. As such, the majority focus on parent chemical clearance through Phase I metabolism. While work is needed to expand the domain of applicability to the broader

use of environmental chemicals, and addition of Phase II metabolism would greatly improve relevance of the models for human risk assessment (Moreau et al., *In Preparation*), predictions of parent chemical clearance can be used in conjunction with IVIVE to estimate oral equivalent doses from *in vitro* bioactivity measures. This approach has been demonstrated recently with estrogenic compounds (Casey et al., 2018). Here, we provide information on user-friendly parent chemical clearance models that are available as freeware or as commercial products or services (see attached worksheet: ADME QSAR).

User-friendly Modeling Platforms for PBPK and IVIVE

Physiologically based pharmacokinetic (PBPK) modeling is a mathematical modeling technique for predicting ADME of synthetic or natural chemical substances in humans and other animal species. These models, which are comprised of a series of differential equations, track the movement of a chemical and its metabolites through the body. PBPK models are traditionally used to predict tissue concentrations from known exposures (forward dosimetry) or to predict external exposures from measured biomarker concentrations (reverse dosimetry). Primers on PBPK modeling and its use in risk assessment can be found elsewhere (Clewell & Clewell et al., 2008).

In vitro to *in vivo* extrapolation (IVIVE) is an application of PBPK modeling that expands upon reverse dosimetry approaches to estimate oral equivalent doses from *in vitro* media concentrations (Figure 7; Rotroff et al., 2010; Wetmore et al., 2012; 2015). IVIVE can range from very high throughput approaches that couple *in vitro* measures of intrinsic clearance and serum binding with assumptions of oral absorption and glomerular filtration (Retroff et al., 2010) to more complex approaches that account for transporter activity and metabolic activation (*i.e.*, quantitative IVIVE - QIVIVE; Yoon et al., 2015). IVIVE is a necessary component of an *in vitro* based risk assessment, in that it converts *in vitro* bioactivity measurements to estimated exposures *in vivo*. Depending on the data that is used to derive a point of departure for a NGR (clinical, epidemiological, *in vitro*, *in silico*) either PBPK modeling or IVIVE will likely be needed to compare external *in vivo* dose to expected exposures. Here, we provide information on user-friendly parent chemical clearance models that are available as freeware or as commercial products or services (see attached worksheet: PBPK IVIVE Modeling). In a Cosmetics Europe funded effort, scientists at ScitoVation previously performed an in-depth evaluation of PBPK platforms, and summarized availability, ease of use, utility/applications in a 2018 Society of Toxicology poster (Efremenko et al., 2018). Here, we used the Efremenko et al. (2018) summary as a starting point and added information from our web search as well as any input from PBPK experts that were interviewed during this project.

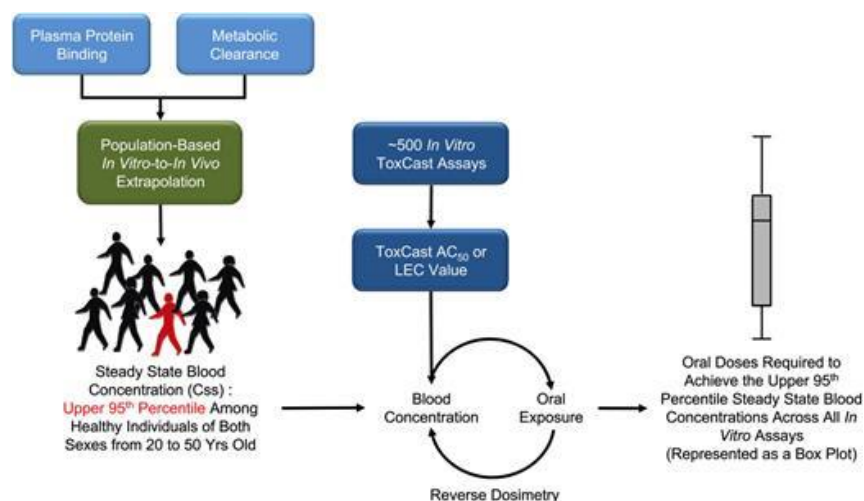


Figure 7. IVIVE Process. Plasma protein binding and metabolic clearance are measured using *in vitro* assays. The pharmacokinetic data were used to parameterize an IVIVE model. Using reverse dosimetry, oral doses were then estimated that would result in a steady-state blood concentration equivalent to the AC₅₀ or LEC value for the *in vitro* assays. Figure reproduced from Wetmore et al., 2012.

Exposure Prediction Models

An effort was made to collect information on exposure models relevant to consumer product exposures. Models vary from lower tier screening models to higher tier models that include more sophisticated analyses of consumer product use. While a number of exposure models exist that can be used for ingredients in consumer products, few accounted for multiple consumer use profiles. The only model identified that allows for customized consumer use scenarios was the Crème family of models (Crème Care™, Crème RIFM™; cremeglobal.com). These models are the result of collaboration with the Research Institute for Fragrance Materials (RIFM) and the cosmetics industry globally (Comisky et al., 2017B; Safford et al., 2017). The collaborative effort includes development of a database with usage data for over 36,000 consumers representing the European and US populations (Comisky et al., 2017B) and exposure models for fragrances (Crème RIFM) and personal care products (Crème Care). The software includes models to account for dermal, ingestion and inhalation exposures, and generates the daily aggregate exposure to each individual in the survey, which allows calculation of a range of population exposure statistics to conduct safety assessments (cremeglobal.com). Other exposure models that were identified are not as directly related to consumer product risk assessment as the Crème models. However, they do have relevance to personal care product exposures and may be useful depending on the risk assessment context (see attached worksheet: Exposure Models). Deterministic, decision tree approaches based on exposure data are also used (Hall et al., 2011), though these have not been reviewed in the current report.

Educational Resources

Initially, searches were performed to identify universities and colleges with undergraduate or graduate degrees in toxicology. Several web-based resources were found that rank college programs based on various criteria, including enrollment, course offerings, and graduation statistics. The top ranked universities for pharmacology and toxicology (often presented as a combined category) are provided in the attached worksheet (see attached worksheet:

University Degree Programs). Included in the worksheet are the universities with master of science, or doctor of philosophy degrees in Pharmacology and/or Toxicology that rank in the Top 200 programs globally according to mastersportal.com and phdportal.com. Additional resources were found that list “top 10” programs for Pharmacology and/or Toxicology in the United States and were also included in the worksheet along with the links to these resources. A preliminary evaluation of the course catalogues was performed for a few universities that were consistently among the highest ranked, to evaluate the extent to which the programs focus on implementation of NAMs in safety decision making. However, in this preliminary analysis, while courses in biochemical or mechanistic toxicology (*i.e., in vitro* toxicology) were common, very few courses were found that had a clear focus on the application of non-animal test methods to risk assessment or practical examples of decision making using non-traditional approaches. As a result, search efforts were refined to search specifically for courses in cosmetic safety testing or courses in 21st century (nonanimal) risk assessment. Among those identified as being at the forefront of efforts to advance the use of non-animal methods in chemical risk assessments are the following:

- Center for Alternatives to Animal Testing (CAAT) at Johns Hopkins University (Maryland, USA) and CAAT-Europe at the University of Konstanz (Germany). CAAT and CAAT-Europe coordinate transatlantic activities to promote the development of new and improved methods in toxicology, to be a partner in strategy development, to provide platforms for different stakeholders, to exchange ideas, and to support the 3R's principle of humane science in different ways. Their goals are the development and promotion of NAMs among the various stakeholders. More information can be found at the CAAT and CAAT-Europe websites (<http://caat.jhsph.edu/programs/index.html> ; <https://www.biologie.uni-konstanz.de/leist/caat-europe/>).
- University of Utrecht. IRAS, the Institute for Risk Assessment Sciences, is an interfaculty research institute within the faculties of Veterinary Medicine, Medicine and Sciences of Utrecht University (Utrecht, Netherlands). The mission of IRAS is to provide education and research on the human health risks of exposure to potentially harmful agents in the environment, at the workplace and through the food chain. The program at IRAS has a strong emphasis on the incorporation of biokinetic (aka, PBPK and IVIVE) models into the risk assessment process. Through collaboration with Utrecht's iLab, students can work with startups toward practical application of non-animal methods to chemical safety decisions. More information can be found at the university website (<https://www.uu.nl/en/organisation/faculty-of-veterinary-medicine/about-the-faculty/departments/iras>; <https://www.utrechtinnovatielab.nl/en/life-sciences-chemistry/about-ilab>).
- Canadian Center for Alternatives to Animal Methods (CCAAM). The CCAAM at the University of Windsor (Ontario, Canada) was launched in 2017 to promote the replacement of animals in Canadian biomedical research, education, and regulatory testing through 21st century science, innovation, and ethics. The activities of the center are focused on three pillars: research, academics, and regulation:
 - Research: “Conduct hypothesis-driven fundamental biomedical research using only human-based biomaterials and human biology-based methodologies”

- Academics: “Establish first-of-kind academic programs in Animal Replacement Science to train the next generation of scientists, ethicists, regulators, and policy makers”
- Regulation: “Expedite the development, validation, and regulatory acceptance of alternative chemical safety testing methods in Canada and contribute to global efforts in a uniquely Canadian way”

More information can be found at the website (<http://www.uwindsor.ca/ccaam/>).

In addition to the brick and mortar universities, we searched for on-line courses and continuing education courses that were focused on the implementation of NAMs into chemical decision-making. While many NAM-focused courses were found, few had the clear objective of implementing NAMs into real-world decision making for chemical safety. Relevant (ongoing) courses are included in the attached worksheet (see attached worksheet: Continuing education).

Of particular relevance to non-animal cosmetics risk assessment is the annual short course “Safety Assessment of Cosmetics in the EU” led by Prof. Vera Rogiers at the Vrije Universiteit Brussel (VUB) (<https://www.safetycourse.eu/index.html#msg-box3-0>). This 1-week course is developed in collaboration with the cosmetic industry. Key issues in safety evaluation of cosmetics are addressed including: application of alternative methods and the interpretation of the results, how and when animal testing can/must be done, toxicological requirements for cosmetic ingredients testing in the EU. The possibility to pass a written exam is also included. The program for the 2019 course can be found at https://www.safetycourse.eu/FINAL_Programme_S2019.pdf.

Discussion

The efforts described in this scoping document are meant to collect information on the resources that are available for education and implementation of non-animal methods in cosmetics testing. This document provides a snapshot of the current technologies that are available to various stakeholders and serves as a starting point for conversations on the next, most important steps to increasing adoption of these methods.

One step may be increasing awareness of the available methods and test providers. The list developed here would serve as valuable resource for these providers and the details of the assays are well described in the cited literature. The challenge will be keeping the list of resources up to date. With the rapid development of computational and *in vitro* methods, this resource will need constant updating in order to stay current.

While the methods are rapidly evolving, the limiting factor to acceptance may actually be an understanding – and a broad consensus – on how to apply the output of these *in vitro* and *in silico* models to a safety decision. Much work has been done within individual personal care product companies – and within consortiums such as Cosmetics Europe and the Personal Care Products Counsel – to develop and implement safety assessment strategies. However, a concerted search of public resources provided few educational resources into these safety assessment strategies. Traditional educational institutions focus mostly on methodologies and research strategies, rather than the safety decision process (particularly with non-animal

methods). Short courses focus on development, validation and acceptance of non-animal methods, but most fall short of defining real world risk assessment processes. Only a single course was identified which clearly fits with goal of practical guidance on applying non-animal methods to cosmetics risk assessment. With the current AFSA collaboration, we have a tremendous opportunity to develop the educational resources to reach a broader audience, including research, industry and regulatory communities, and to promote the implementation of these non-animal safety assessments.

References

- Ames BN, Gurney EG, Miller JA, Bartsch H. 1972. Carcinogens as frameshift mutagens: metabolites and derivatives of 2-acetylaminofluorene and other aromatic amine carcinogens. *Proceedings of the National Academy of Sciences of the United States of America*. 69 (11): 3128–32.
- Api AM, Belsito D, Bruze M, Cadby P, Calow P, Dagli ML, Dekant W, Ellis G, Fryer AD, Fukayama M, Griem P, Hickey C, Kromidas L, Lalko JF, Liebler DC, Miyachi Y, Politano VT, Renskers K, Ritacco G, Salvito D, Schultz TW, Sipes IG, Smith B, Vitale D, Wilcox DK. 2015. Criteria for the Research Institute for Fragrance Materials, Inc. (RIFM) safety evaluation process for fragrance ingredients. *Food Chem Toxicol*. 82:Suppl:S1-S19.
- Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. 2015. Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model. *Environ Sci Technol*. 49(14):8804-14.
- Burgdorf T, Piersma AH, Landsiedel R, Clewell R, Kleinstreuer N, Oelgeschläger M, Desprez B, Kienhuis A, Bos P, de Vries R, de Wit L, Seidle T, Scheel J, Schönfelder G, van Benthem J, Vinggaard AM, Eskes C, Ezendam J. 2019. Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory toxicology - Evolution versus revolution. *Toxicol In vitro*. 59:1-11.
- Casati, S., Aschberger, K., Barroso, J. et al. 2018. Standardisation of defined approaches for skin sensitisation testing to support regulatory use and international adoption: position of the International Cooperation on Alternative Test Methods. *Arch Toxicol*. 92: 611.
- Clewell RA, Clewell HJ 3rd. 2008. Development and specification of physiologically based pharmacokinetic models for use in risk assessment. *Regul Toxicol Pharmacol*. 50(1):129-43.
- Comiskey D, Api AM, Barratt C, Daly EJ, Ellis G, McNamara C, O'Mahony C, Robison SH, Safford B, Smith B, Tozer S. 2017A. Novel database for exposure to fragrance ingredients in cosmetics and personal care products. *Regul Toxicol Pharmacol*. 72(3):660-72.
- Comiskey D, Api AM, Barrett C, Ellis G, McNamara C, O'Mahony C, Robison SH, Rose J, Safford B, Smith B, Tozer S. 2017B. Integrating habits and practices data for soaps, cosmetics and air care products into an existing aggregate exposure model. *Regul Toxicol Pharmacol*. 88:144-156.

Dent M, Amaral RT, Da Silva PD, Ansell J, Boislevé F, Hatao M, Hirose A, Kasai Y, Kern P, Kreiling R, Milstein S, Montemayor B, Oliveira J, Richarz A, Taalman R, Vaillaancourt E, Verma R, Posada NVOC, Weiss C, Kojima H. 2018. Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. *Comput Toxicol.* 7:20-26.

ECHA 2016A. ECHA Practical guide – How to use alternatives to animal testing to fulfil your information requirements for REACH registration. Version 2.0 – July 2016.

ECHA 2016B. ECHA Practical guide – How to use and report (Q)SARs version 3.1 – July 2016.

Efremenko AY, Nicola CI, Hack E, Ellison C, Mansouri K, McMullen PD, Klaric M, Hewitt N, Clewell. 2018. Evaluation of Platforms Used for Fit-for-Purpose PBPK Modeling in Chemical Safety Assessment. *The Toxicologist, Supplement to Toxicological Sciences*, 150 (1), Abstract #3169.

Ellison CA, Blackburn KL, Carmichael PL, Clewell HJ 3rd, Cronin MTD, Desprez B, Escher SE, Ferguson SS, Grégoire S, Hewitt NJ, Hollnagel HM, Klaric M, Patel A, Salhi S, Schepky A, Schmitt BG, Wambaugh JF, 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. *Regul Toxicol Pharmacol.* 103:63-72.

Embry MR, Bachman AN, Bell DR, Boobis AR, Cohen SM, Dellarco M, Dewhurst IC, Doerrner NG, Hines RN, Moretto A, Pastoor TP, Phillips RD, Rowlands JC, Tanir JY, Wolf DC, Doe JE. 2014. Risk assessment in the 21st century: roadmap and matrix. *Crit Rev Toxicol.* 44 Suppl 3:6-16.

Gocht T, Berggren E, Ahr HJ, Cotgreave I, Cronin MT, Daston G, Hardy B, Heinzle E, Hescheler J, Knight DJ, Mahony C, Peschanski M, Schwarz M, Thomas RS, Verfaillie C, White A, Whelan M. 2015. The SEURAT-1 approach towards animal free human safety assessment. *ALTEX.* 32(1):9-24.

Hall B, Steiling W, Safford B, Coroama M, Tozer S, Firmani C, McNamara C, Gibney M. 2011. European consumer exposure to cosmetic products, a framework for conducting population exposure assessments Part 2. *Food Chem Toxicol.* 49(2):408-22.

Judson R, Kavlock R, Martin M, Reif D, Houck K, Knudsen T, Richard A, Tice RR, Whelan M, Xia M, Huang R, Austin C, Daston G, Hartung T, Fowle JR 3rd, Wooge W, Tong W, Dix D. 2013. Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX.* 30(1):51-6.

Kleinstreuer NC, Karmaus A, Mansouri K, Allen DG, Fitzpatrick JM, Patlewicz G. 2018. Predictive Models for Acute Oral Systemic Toxicity: A Workshop to Bridge the Gap from Research to Regulation. *Comput Toxicol.* 8(11):21-24.

Kroes R, Kleiner J, Renwick A. 2005. The Threshold of Toxicological Concern Concept in Risk Assessment. *Toxicol Sci.* 86(2):226–230.

Kroes R, Renwick AG, Feron V, Galli CL, Gibney M, Greim H, Guy RH, Lhuguenot JC, van de Sandt JJ. 2007. Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. *Food Chem Toxicol.* 45(12):2533-62.

Manganelli S, Roncaglioni A, Mansouri K, Judson RS, Benfenati E, Manganaro A, Ruiz P. 2019. Development, validation and integration of *in silico* models to identify androgen active chemicals. *Chemosphere.* 220:204-215.

Mansouri K, Abdelaziz A, Rybacka A, Roncaglioni A, Tropsha A, Varnek A, Zakharov A, Worth A, Richard AM, Grulke CM, Trisciuzzi D, Fourches D, Horvath D, Benfenati E, Muratov E, Wedebye EB, Grisoni F, Mangiatordi GF, Incisivo GM, Hong H, Ng HW, Tetko IV, Balabin I, Kancherla J, Shen J, Burton J, Nicklaus M, Cassotti M, Nikolov NG, Nicolotti O, Andersson PL, Zang Q, Politi R, Beger RD, Todeschini R, Huang R, Farag S, Rosenberg SA, Slavov S, Hu X, Judson RS. 2016. CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. *Environ Health Perspect.* 124(7):1023-33.

Mansouri K, Grulke CM, Judson RS, Williams AJ. 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. *J Cheminform.* 10(1):10.

Moreau M, Mallick P, Smeltz M, Pendse S, Leonard J, McMullen P, Phillips MB, Clewell R, Clewell HJ, Yoon M. Incorporation of metabolite exposure in high-throughput *in vitro* to *in vivo* extrapolation (HT-IVIVE). *In preparation.*

OECD. 2017A. Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Serious Eye Damage and Eye Irritation. Series on Testing & Assessment, No. 263. ENV/JM/MONO(2017)15.

OECD. 2017B. New guidance document on an integrated approach on testing and assessment (IATA) for skin corrosion and irritation. Series on Testing and Assessment, No. 203. ENV/JM/MONO(2014)19.

Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, Lecluyse EL, Andersen ME, Judson RS, Smith CM, Sochaski MA, Kavlock RJ, Boellmann F, Martin MT,

Roy S, Kulkarni R, Hewitt NJ, Aardema MJ. 2016. The EpiDerm™ 3D human reconstructed skin micronucleus (RSMN) assay: Historical control data and proof of principle studies for mechanistic assay adaptations. *Mutat Res Genet Toxicol Environ Mutagen.* 805:25-37.

Reif DM, Wambaugh JF, Thomas RS. 2010. Incorporating human dosimetry and exposure into high-throughput *in vitro* toxicity screening. *Toxicol Sci.* 117(2):348-58.

Safford RJ. 2008. The Dermal Sensitisation Threshold- a TTC approach for allergic contact dermatitis. *Regul Toxicol Pharmacol.* 51(2):195-200.

Safford B, Api AM, Barratt C, Comiskey D, Ellis G, McNamara C, O'Mahony C, Robison S, Rose J, Smith B, Tozer S. 2017. Application of the expanded Creme RIFM consumer exposure model to

fragrance ingredients in cosmetic, personal care and air care products. Regul Toxicol Pharmacol. 86:148-156.

Schüürmann G, Ebert RU, Tluczkiewicz I, Escher SE, Kühne R. 2016. Inhalation threshold of toxicological concern (TTC) - Structural alerts discriminate high from low repeated-dose inhalation toxicity. Environ Int. 88:123-132.

Sjöström M, Kolman A, Clemenson C, Clothier R. 2008. Estimation of human blood LC50 values for use in modeling of *in vitro-in vivo* data of the ACuteTox project. Toxicol In vitro. 22(5):1405-11.

USEPA 2016. Toxic Release Inventory – Chemical Hazard Information Profiles (TRI-CHIP). (Nov 17, 2016). <https://www.epa.gov/toxics-release-inventory-tri-program/download-tri-chip>.

USEPA 2018. TSCA chemical substance inventory. <https://www.epa.gov/tsca-inventory>.

Wetmore BA, Wambaugh JF, Allen B, Ferguson SS, Sochaski MA, Setzer RW, Houck KA, Strobe CL, Cantwell K, Judson RS, LeCluyse E, Clewell HJ, Thomas RS, Andersen ME. 2015. Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted *In vitro* Bioactivity to Inform Chemical Toxicity Testing. Toxicol Sci. 148(1):121-36.

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, Clewell HJ 3rd, Dix DJ, Andersen ME, Houck KA, Allen B, Judson RS, Singh R, Kavlock RJ, Richard AM, Thomas RS. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci. 125(1):157-74.

Whelan M, Eskes C. 2016. Evolving the Principles and Practice of Validation for New Alternative Approaches to Toxicity Testing. Adv Exp Med Biol. 856:387-399.

Williams FM, Rothe H, Barrett G, Chiodini A, Whyte J, Cronin MT, Monteiro-Riviere NA, Plautz J, Roper C, Westerhout J, Yang C, Guy RH. 2016. Assessing the safety of cosmetic chemicals: Consideration of a flux decision tree to predict dermally delivered systemic dose for comparison with oral TTC (Threshold of Toxicological Concern). Reg toxicol pharmacol. 76:174-86.

Wilson SL, Ahearne M, Hopkinson A. 2015. An overview of current techniques for ocular toxicity testing. Toxicology. 327:32-46.

Yang C, Barlow SM, Muldoon Jacobs KL, Vitcheva V, Boobis AR, Felter SP, Arvidson KB, Keller D, Cronin MTD, Enoch S, Worth A, Hollnagel HM. 2017. Thresholds of Toxicological Concern for cosmetics-related substances: New database, thresholds, and enrichment of chemical space. Food Chem Toxicol. 109(Pt 1):170-193.

Yoon M, Blaauw BJ, Clewell HJ. 2015. Quantitative *in vitro* to *in vivo* extrapolation (QIVIVE): An essential element for *in vitro*-based risk assessment. Toxicology. 332:1-3