REPORT FOR THE INTERNATIONAL COOPERATION ON COSMETICS REGULATION

Joint Regulators-Industry Working Group: **Integrated Strategies for Safety Assessments of Cosmetic Ingredients - Part I**

Report is:

☐ Prepared for ICCR ☑ ICCR Report ☐ ICCR Guidance

Authors:

Renata Amaral (1), Pedro Amores Da Silva (1), Jay Ansell (2), Fanny Boisleve (3), Matt Dent (4)#, Masato Hatao (5), Akihiko Hirose (6), Yutaka Kasai (7), Hajime Kojima (6)#, Petra Kern (8), Reinhard Kreiling (9), Stanley Milstein (10), Julcemara Oliveira (11), Andrea Richarz (12), Rob Taalman (13), Eric Vaillancourt (14), Rajeshwar Verma (10), Nashira Campos Vieira (11), Craig Weiss (15)

1) Brazilian Association of the Cosmetic, Toiletry and Fragrance Industry (ABIPHEC), Av. Paulista, 1313 Cerqueira César, São Paulo, SP 01311-000, Brazil
2) US Personal Care Products Council (PCPC), 1620 L St. NW, Suite 1200, Washington, D.C. 20036, USA
3) Johnson & Johnson Santé Beauté France, Domaine de Maigremont, CS 10615, F-27106 VAL DE REUIL Cedex, France
4) Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedfordshire MK44 1LQ, UK (Co-Chair)
5) Japan Cosmetic Industry Association (JCIA), Metro City Kamiyacho 6F, 5-1-5, Toranomon, Minato-ku, Tokyo, 105-0001 Japan
6) National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, 158-8501, Tokyo, Japan
7) Kao corporation, External Relations & Government Affairs 2-1-3, Bunka, Sumida-Ku, Tokyo, 131-8501 Japan
8) Proctor and Gamble Services Company NV, Temselaan 100, B-1853 Strombeek-Bever, Belgium
9) Clariant Produkte (DE) GmbH, Global Toxicology and Ecotoxicology, Am Unisys-Park 1, 65843 Sulzbach, Germany
10) US Food and Drug Administration (US FDA), Office of Cosmetics and Colors, Center for Food Safety and Applied Nutrition, 5001 Campus Way, College Park, MD 20740, USA
11) Brazilian Health Regulatory Agency (ANVISA), Gerência de Cosméticos, SIA Trecho 5, lote 200, Area Especial 57 – CEP 71205-050
12) European Commission, Joint Research Centre (JRC), Directorate for Health, Consumers and Reference Materials, Chemical Safety and Alternative Methods Unit, Via E. Fermi 2749, 21027 Ispra (VA), Italy
13) Cosmetics Europe, Avenue Herrmann-Debroux 40, 1160 Auderghem, Belgium
14) Health Canada (HC), Consumer Product Safety Directorate, Healthy Environments and Consumer Safety Branch, 269 Laurier Ave. W., Ottawa, ON, K1A 0K9, Canada
15) Independent Cosmetic Manufacturing and Distributors (ICMAD), 21925 Field Parkway, Suite 2015, Deer Park, IL 60010, USA

Co-chair of JWG

Date of Preparation: June, 6, 2017	**Status:** FINAL DELIVERABLE

Table of Contents

1. Purpose

The International Cooperation on Cosmetics Regulation (ICCR) held its ninth annual meeting (ICCR-9) November 5, 2015 in Brussels, Belgium[1]. At this meeting, discussions related to the evolution of the working group on *in silico* methods/quantitative structure activity relationships (QSARs) and updates on alternatives to animal testing approaches took place. These led to the proposal that a single new joint working group, covering a holistic approach to identify modern methods and Integrated Approaches to Testing and Assessment (IATA), relevant to the safety assessments of ingredients used in cosmetics, be formed.

The purpose of the new ad hoc Joint Regulators-Industry Working Group (JWG) is to outline principles that underpin the integration of novel methods and data in an exposure led approach for the safety assessment of cosmetic ingredients.

2. Scope

The scope of the novel methods that were considered included *in silico* methods (including (Q)SARs and other computational modelling approaches), *in chemico* methods and *in vitro* tests.

It is understood that the science in these areas is rapidly evolving, with a large number of models and approaches referenced in the literature, ranging from exploratory to those that can be considered mature and well developed but whose applicability domain is outside of cosmetic ingredients.

In this report, the JWG focused on the principles that underpin the integration of novel methods and data in an exposure-led approach for the safety assessment of cosmetic ingredients. New approach methodology (NAM)will be addressed in a future report.

3. Selected Abbreviations

AUC	Area Under Curve
ECHA. . . .	European Chemicals Agency
GIVIMP. . . .	Good *in vitro* methods practice
GLP. . . .	Good laboratory practice
HTS. . . .	High Throughput Screening
ICCR. . . .	International Cooperation on Cosmetics Regulation
IATA. . . .	Integrated Approaches to Testing and Assessment
JWG. . . .	Ad hoc Joint Regulators-Industry Working Group
NAM. . . .	New approach methodology
NAS	National Academies of Science
NO(A)EL. . . .	No observed (adverse) effect level
MoA. . . .	Mode of Action
OECD	Organisation for Economic Co-operation and Development
(Q)IVIVE. . . .	(Quantitative) *in vitro* to *in vivo* extrapolation

[1] A more comprehensive discussion of the outcomes from this and previous meetings may be found at the ICCR web site at: http://www.iccrnet.org/chairmanships/

(Q)SAR. . . .	(Quantitative) structure activity relationship
SOP. . . .	Standard operating procedure
SR.	Systematic Review
TTC. . . .	Threshold of Toxicological Concern

4. Introduction

Cosmetic products and ingredients should be safe for consumers under their conditions of use. Historically the safety assessment for some toxicological endpoints relied on animal testing. However, concern for animal welfare, regulatory action and a desire by companies to bring safe products to market without the use of animal testing using more human-relevant data has brought the need for a different approach to evaluating safety. In 2007 the US National Academies of Science (NAS) published a seminal document entitled *Toxicity Testing in the 21st Century, A Vision and a Strategy (*NAS 2007, Krewski *et al., 2010).* This NAS report called for a transformation in toxicity testing, "from a system based on whole-animal testing to one founded primarily on *in vitro* methods that evaluate changes in biological processes using cells...of human origin." This transformation will require the use of new types of data that have not routinely been used in cosmetic safety evaluation. In 2017, the NAS followed up on the conceptual frameworks laid out in the 2007 report and a 2012 report on *Exposure Science in the 21st Century* (NAS, 2012) with the report on *Using 21st Century Science to Improve Risk-Related Evaluations* (NAS, 2017). This new report discusses the advances and challenges in risk assessment related to interpreting and integrating new types (and volumes) of data, with an emphasis on exposure considerations. In parallel, the use of data and information from NAMs has also been discussed in a broader context in Europe in a dedicated European Chemicals Agency (ECHA) Topical Scientific Workshop held in April 2016, identifying their potential and existing barriers to support regulatory decisions for the assessment of chemical substances (ECHA, 2016).

ICCR has produced recommendations for the use of alternatives to animal test methods in cosmetics safety evaluation for a number of years. Given the rapid evolution in the science of toxicological safety assessment, and the opportunities provided by NAMs as described in the above NAS and ECHA reports, a fundamental change in our approach to safety evaluation is becoming possible. The purpose of this ICCR report, prepared by an *ad hoc* Joint Regulators-Industry Working Group (JWG) is therefore to outline principles that underpin the integration of novel methods and data in an exposure-led approach for the safety assessment of cosmetic ingredients. This builds on a previous report for ICCR which describes the overall principles of cosmetic product safety evaluation.

5. Discussion

This report summarizes major overarching principles for incorporating NAMs into an integrated strategy for risk assessment of cosmetics ingredients (or 'Next Generation' risk assessment), along with examples showing their usefulness to safety evaluation.

In the context of cosmetics safety evaluation, a Next Generation risk assessment is defined as an exposure-led, hypothesis driven risk assessment approach that incorporates one or more NAMs to ensure that use of a cosmetic product does not cause harm to consumers.

The 4 main overriding principles are articulated in this definition:

1. The overall goal is a human safety risk assessment
2. The assessment is exposure led
3. The assessment is hypothesis driven
4. The assessment is designed to prevent harm (i.e. distinguish between adaptation and adversity)

The following 3 principles describe how a Next Generation risk assessment should be conducted:

5. Using a tiered and iterative approach
6. Following an appropriate appraisal of existing information
7. Using robust and relevant methods and strategies

Finally, 2 principles for documenting Next Generation risk assessments are described:

8. The logic of the approach should be transparently and explicitly documented
9. Sources of uncertainty should be characterized and documented

This report is intended to help those involved in cosmetic safety assessment build integrated safety assessments without generating animal data.

6. Principles underpinning the use of NAMs in the risk assessment of cosmetic ingredients

Principle 1: The overall goal is a human safety risk assessment

Firstly, the safety assessment should enable a decision to be made on the safety of the ingredient/product to humans, not be designed as a battery of tests to replicate the results of animal studies.

Each of the cosmetics regulatory authorities in ICCR (i.e., Brazil, Canada, the European Union, Japan and the United States) has an established framework governing cosmetic safety. While a side by side comparison will show some differences across regions the commonalties are far more numerous. These includes uniform agreement that pre-market approval is not necessary for the vast majority of products; ingredients should be labeled; good manufacturing practices should be respected; and as an undeviating cornerstone -- the responsibility of Manufacturers to substantiate the safety of the cosmetic product.

Thus, each region has as its overarching principle that cosmetics must be safe. Similarly, it is consistent across all five regions that the regulations do not prescribe specifically how the safety of the cosmetic must be determined. No iterative list of tests is required but rather it is the responsibility of the manufacture to assure that cosmetics placed in the market are safe for the consumer. In this regard cosmetics are unlike many other regulated product categories where a fixed testing data set, often including specified animal tests, is obligated by law or regulation. When considered in light of the move away from animal testing in general, this presents cosmetic manufacturers with opportunities to ensure safety risk assessments are grounded in human biology rather than replicating the results of a prescriptive list of animal tests. This is especially important for mechanism-based risk assessments that are informed by changes in cellular signalling pathways in cells or tissues of human origin, because results of such tests cannot (and should not) be 'validated' against the results of apical animal studies.

Principle 2: The assessment is exposure led

Exposure assessment is *"the process of estimating or measuring the magnitude, frequency, and duration of exposure to an agent, along with the number and characteristics of the population exposed. Ideally, it describes the sources, routes, pathways, and uncertainty in the assessment"* (IPCS, 2004).

Exposure assessment is one of the four essential steps in cosmetic ingredient human safety assessment; others include hazard identification, dose-response assessment, and risk characterization. While historically, the safety assessment was hazard driven, it has now shifted towards exposure driven approaches (SCHER *et al.,* 2013). Estimating human exposure as early as possible in the safety assessment is crucial. This is because in an exposure-driven paradigm exposure estimates will define the degree of hazard data needs and guide further data generation. For example, techniques such as exposure-based waiving using thresholds of toxicological concern (TTC) may be sufficient to assure the safety in case of very low

exposures, and calculated internal exposure concentrations will guide concentrations to be used for possible *in vitro* tests performed for the risk assessment and may also contribute to the mode of action hypothesis by indicating target organs. As well as the cosmetic ingredient itself, it may also be necessary to characterize exposure to any relevant impurities present in the ingredient and/or metabolites.

Exposure may be estimated using an iterative tiered approach, ranging from screening-level to a refined exposure assessment which considers both external (applied dose) and internal (systemic) exposure. Exposure data feeding into a human safety assessment can be deterministic or probabilistic, and describe exposure from a single product or combined exposure from multiple products or sources (aggregate exposure). As the first step, a screening level assessment using basic tools (e.g., simple exposure calculations, default values, rules of thumb, conservative assumptions, deterministic approaches) can be conducted. Depending on the results of the screening-level assessment further evaluation through refinements of the input data and exposure assumptions or by using more advanced models, such as skin absorption studies and probabilistic exposure assessments may be warranted. Probabilistic models rely on data distributions instead of point values and hence result in exposure distributions better characterizing realistic consumer exposures.

In 'traditional' cosmetic safety evaluations exposure is often expressed as either the applied dose per unit area (e.g. $\mu g/cm^2$ for local effects) or as total body burden (e.g. mg/kg body weight/day for systemic effects). For risk assessments that integrate NAM, depending on the methodology and health effect being evaluated it is likely that exposure will be expressed on an internal basis using metrics such as Cmax or area under curve (AUC), and calculated using relevant pharmacokinetic models. In the cases where a quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) is required, the QIVIVE must allow a valid comparison between the actual dose in the *in vitro* test system rather than the applied dose. In these cases the free concentration of the test chemical may be a more valid metric than the total applied dose (Groothuis *et al.,* 2015)

At each level of assessment one needs to decide if the degree of confidence in the data is good enough to achieve the purpose of the assessment or if successive iterations using more data or refinements are required. The ability to integrate the refined exposure assessment with the hazard identification and dose-response assessment into the human safety assessment (and possibly incorporate this into regulatory decision making) is influenced by the quality of the exposure characterization.

Principle 3: The assessment is hypothesis driven

Historically, safety assessments were animal-based relying on the assumption that clinical and pathological effects seen at high doses in animal models are relevant and titratable to much lower exposures (often by a different exposure route) in humans. More specific mode of action (MoA) hypotheses only tend to be articulated once adverse effects are seen in intact animals and it is then determined whether the effects are relevant to humans.

In the context of non-animal toxicological safety assessments it is important to use appropriate information to establish a hypothesis (or hypotheses) about the biologically relevant MoA that may be associated with a specific chemical exposure.

The Oxford Dictionary defines hypothesis as:

"*A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation*".

Furthermore, from a statistical perspective, a (null) hypothesis is an *affirmation* that is testable using appropriate statistical analysis.

The 'limited evidence' that could be used to establish the hypothesis should include all available existing data. This could include *in vitro* or *in vivo* data, read across and *in silico* predictions. This being said, care needs to be taken not to bias the hypothesis based on the focus of previous investigations. For example, if a chemical has been researched and shown to interact with a specific receptor, this should not be the entire focus of the safety evaluation as other important MoA could be missed. This is where a broad high throughput screen (HTS) could be used to inform potential MoA, to consider alongside the existing data. The assays represented in this HTS could include consideration of stress response pathways (e.g. oxidative stress) as well as specific protein/receptor interactions (e.g. oestrogen receptor activity). Any available animal data should be used with care at this step of the safety evaluation. If relevant *in vivo* data are sufficient to complete a risk assessment using traditional methodologies, the expectation is that a risk assessment based on the *in vivo* data will be performed. However, where there are significant data gaps in the existing *in vivo* dataset, existing animal data should only be used if it can help to identify potential modes of action which could help to establish or refine the hypothesis. Furthermore, in this instance, as the goal is to produce a human-relevant safety assessment, it is important that the hypothesis is not focussed on predicting or confirming reported adverse effects in the available limited animal data. For example, if limited animal test data suggest hepatic toxicity at a particular dose, this information is only of use if it can be used alongside the HTS data help identify the MoA that may cause adverse effects in humans. The hypothesis in this example should therefore not be 'Chemical X causes liver toxicity in rats after an oral dose of 10 mg/kg/day', because this will send the safety assessment on course to evaluate changes in animal models that may not be relevant for humans. Rather, the hypothesis established should be focussed on the MoA thought to be responsible. Examples of such hypotheses based on HTS screens could be:

'At relevant exposures, Chemical X perturbs the p53 pathway which results in increased cancer risk in consumers', or 'At relevant exposures, Chemical X does not cause adverse effects in consumers due to an ability to antagonize the androgen receptor'.

Conversely, following the assessment of the available data and the HTS screen, the hypothesis could be 'At relevant exposures the biological activity of Chemical X is insufficient to cause adverse effects in consumers'. Similarly, this hypothesis may require refinement, for example by repeating the HTS in different cell lines to increase confidence in the safety evaluation, and

it should be tested using appropriate statistical analysis while all underlying assumptions should be clearly defined.

Determining the appropriate hypothesis (or hypotheses) will enable identification of relevant questions that need to be answered using appropriate techniques to complete a safety evaluation.

Principle 4: The assessment is designed to prevent harm

It is normal practice in the interpretation of animal toxicity studies to distinguish between adaptive effects of treatment (effect level) and those that are considered adverse (adverse effect level), with the point of departure used in risk characterization being based on the dose expected to cause no or minimal adverse effect (no observed adverse effect level -:NOAEL). A limitation of this approach is that the biological mechanisms that underlie these adverse effects are rarely known. In contrast, most NAMs are based on defining a chemical's biological activity to inform a mechanism-based risk assessment. When used in isolation, many NAMs are not designed to distinguish between a biological effect (a treatment-related change detectable in the test system) and an adverse effect (an effect that will result in an adverse health effect in humans).

Whilst it may be relatively straightforward to identify an *in vitro* concentration that results in perturbation of e.g. a stress response pathway by measuring altered levels of signalling molecules or the expression of genes controlling the pathway of interest, determining a dose which could result in adverse health effects in humans is much more of a challenge. An important reason for this is that the homeostatic responses that allow an integrated *in vivo* system to compensate for stress are missing in isolated *in vitro* test systems. Where no biological effects at all are predicted to occur at human-relevant exposures this is not an issue, because if there are no effects there can be no adversity. However, many NAMs can identify biological effects with great sensitivity, meaning that in many cases it will be necessary to develop tools and approaches to enable experimenters and risk assessors to distinguish between a dose that may result in an adaptive or adverse response.

Although this could seem like an ambition that is currently out of reach, this may be accomplished using pragmatic approaches such as benchmarking exposure and effect concentrations against different chemicals with similar MoA where there is a strong presumption of safety (or otherwise). The use of more elaborate approaches such as advanced *in vitro* systems (e.g. 3D models) which are more *in vivo* like, or bespoke computational models capable of modelling the dynamics of the *in vivo* system are additional tools to refine the risk assessment. Whichever approach is taken Principle 5 (using a tiered and iterative process) should guide the process so that the level of work is proportional to the level of concern, ensuring that work stops once there is enough precision to make a decision (Embry *et al.,* 2014).

Principle 5: Using a tiered and iterative approach

The amount of resources allocated to conducting a risk assessment should be based on, and be proportional to, the level of concern. Because resources are finite, the greatest amount of money, time and effort should be assigned to the most potentially significant risks.
Several factors can guide the resource prioritization process. Such factors include, but are not limited to, the level of severity of the potential injury, the level of exposure involved, whether a vulnerable population (for example, children, pregnant women, seniors) can be identified, whether the hazard remains present when the cosmetic product is used in accordance with its intended use, or the level of refinement of the hypothesis to be tested.

To ensure the allocated resources allow gathering the optimal level of precision to make a decision, it may be useful to use a tiered approach for risk assessment which would thereby involve tiered approaches for toxicity and exposure estimation. In some circumstances, it may be sufficient to develop low tier estimates based on (Q)SARs paired with a Threshold of Toxicological Concern (TTC) approach using minimal information such as physico-chemical properties (Kroes *et al*., 2004; Embry *et al*., 2014). As an exposure-based waiving approach, the TTC has been found to be broadly applicable to cosmetics (Kroes *et al*., 2007, Worth *et al*., 2012, SCCS, 2016, Williams *et al.,* 2016) and can also be used for inhalation exposure to aerosol ingredients (Carthew *et al*., 2009). If low tier estimates yield enough information to make a decision, then there is no need to allocate further resources to obtain higher-tier estimates. If more refinement is required, however, increasing resources could be assigned to produce higher tier *in vitro* estimates involving predictive assays paired with IVIVE using deterministic exposure models encompassing population-specific exposures, or further refinement yet using probabilistic exposure scenarios. Finally, if an even-greater level of refinement is needed, further resources can be allotted to produce estimates based on dose-response for the relevant MoA combined with biomonitoring data (Embry *et al*., 2014).

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.

Principle 6: Following an appropriate appraisal of all existing information

It is important to ensure that all available relevant knowledge and information is used to shape the scope and direction of the assessment. It is recommended to use systematic review methodology to identify, select and critically appraise relevant information to ensure that all the steps of the risk assessment process (hazard identification, hazard characterization, exposure assessment, risk estimate) are based on relevant and robust data. Furthermore, the findings of systematic reviews could provide information as input into risk assessment models.

A systematic review (SR) is an overview of existing evidence pertinent to a clearly formulated question, which uses pre-specified and standardized methods to identify and critically appraise relevant research, and to collect, report and analyze data from the studies that are included in the review (EFSA, 2010). Statistical methods to synthesize the results of the included studies (meta-analysis) may or may not be used in the process. Due to their

methodological rigour, transparency and reproducibility SRs are different from narrative reviews and can provide several important functions in the risk assessment process.

Most simply, an effective SR will prevent the conduct of redundant experiments not necessary to complete the risk assessment, and reduce bias in the evaluation of existing information. Where new data needs are identified, SRs may improve the design and therefore the relevance and reliability of new experiments.

Secondly, a SR can allow the use of both high and low quality data. Where the evidence located is of high quality the review may be able to produce an estimate of effect that is unbiased and more precise than those available from any individual study. If the research located is of poor quality then the review will document the limitations and flaws with the existing evidence, formally identify knowledge gaps, and make informed proposals for the weight given to the data in the overall assessment.

Principle 7: Using robust and relevant methods and strategies

For confidence in the validity of the safety assessment, it should be based on robust and relevant methods.

Criteria to assess this may include adherence to the Organisation for Economic Co-operation and Development (OECD) Test Guidelines and work in a relevant quality system (e.g. good laboratory practice (GLP)), applying both to performing *in vitro* assays and evaluation of the quality of existing study data. In addition, Good In Vitro Methods Practice (GIVIMP) has been introduced in order to reduce variability in *in vitro* methods for regulatory safety assessment and to allow harmonisation of approaches (Coecke *et al.,* 2016, OECD, 2016a). GIVIMP is based on good scientific and good quality practices, including considerations on standard operating procedures (SOPs) of *in vitro* methods, the minimum SOP requirements and reporting features necessary as well as describing good experimental design and establishing acceptance criteria for *in vitro* methods and performance standards. Similarly, any *in silico* methods used should be sufficiently documented, transparent and reproducible (see also Principle 8).

However, it should be noted that new approaches need not necessarily be formally validated, endorsed by regulatory authorities, or performed to GLP to be useful. In the ECHA Topical Scientific Workshop on New Approach Methodologies in Regulatory Science held in April 2016, the usefulness of NAM for a number of regulatory uses was stressed, especially in providing pertinent information about MoA (ECHA, 2016). In 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.

The interpretation and combination of information from different methods to inform the risk assessment can be standardised in defined approaches (DA) to testing and assessment, which can be components of IATA. DA are rule-based approaches. Data generated by different methods (*in silico*, *in chemico*, *in vitro*, *in vivo*), which are deemed relevant and fit for purpose

for the health effect considered, are evaluated using a fixed data interpretation procedure (OECD, 2016b). A DA can have the form of a sequential testing strategy or an integrated testing strategy. An example is the guidance on reporting of DA for use within IATA for skin sensitisation (OECD, 2016c). Any remaining uncertainties relating either to the methods used or to the risk assessment strategy should be transparently documented (see Principle 9).

Principle 8: The logic of the approach should be transparently and explicitly documented

When conducting a risk assessment, all data used, assumptions, methodology and software should be clearly documented and be available for an independent review. More specifically, the following should be considered: The problem formulation, the assumption(s), the rationale for each assumption, the hypothesis(es), the potential MoA, and why the selected approach is valid should all be clearly articulated. Hyperlinks (preferably direct object identifiers) to freely accessible peer-reviewed literature should be provided along with the original risk assessment; the methods, reagents, cells, tissues, and statistical tests (including outlier treatment) should be detailed and unambiguous; for *in silico* methods, it should be stated whether commercial or open-source application software is used. Such software should be high-quality (including the statistical level of confidence in the predictions and the determination of the domain of applicability and associated with transparent software descriptions and processes to generate the predictions. Because software may be available in numerous versions, it is important to document the exact version used and, if possible, the substances used to build the model (the training set) to ensure replicability and relevance. For documentation of QSARs, the QSAR Model Reporting Format could be used, which follows the OECD principles for validation of QSARs (OECD, 2007)

The levels of transparency and clarity need to be such as to allow a non-expert, decision-making reviewer to understand the data and reasoning of the assessment, to start an independent review from the beginning, and to reach the same conclusions as those outlined in the original analysis.

Principle 9: Sources of uncertainty should be characterized and documented

"Uncertainty can be caused by limitations in knowledge (e.g. limited availability of empirical information), as well as biases or imperfections in the instruments, models or techniques used" (ECHA, 2012). There are limitations, biases or imperfections leading to uncertainty in any risk assessment regardless of the methodology used. Traditional (animal-based) risk assessments have evolved strategies to deal with uncertainty. These include development of regulatory guidance describing data needed to complete a risk assessment, test guidelines to describe how studies should be performed, and guidance documents describing how data should be interpreted. In terms of safety decision making some uncertainties are addressed with the use of default or data-driven uncertainty factors (Renwick and Lazarus, 1998). These uncertainty factors (also referred to as safety or assessment factors) are intended to allow for possible inter-species and inter-individual differences in response to test chemicals (both

tokicokinetics and tokicodynamics) as well as other considerations such as duration of study and overall quality of the database. Therefore, although not always explicit, uncertainty has always been a feature of toxicological risk assessment and has been addressed in a variety of ways.

All sources of uncertainties should be identified and characterised to provide transparency for the decision process. Variability and uncertainty should be distinguished and all different sources should be considered, e.g. measurement or method uncertainties (EFSA, 2016).
Where novel tools are used in the safety or risk assessment process, especially where guidance for the evaluation of these approaches is not available, the uncertainty associated with their use should be explicitly described, also considering that the results from different types of methods will be integrated in a weight of evidence approach. For example, where data from *in vitro* studies are directly used in a risk assessment, sources of uncertainty could include how representative the test system is of human cells/tissues, i.e. the mechanistic and human relevance. Rather than relying on conservative default factors to address this uncertainty, it would be scientifically more robust to transparently characterize this uncertainty and where required develop a strategy to reduce the uncertainty, e.g. by generating data addressing limitations in knowledge. The data quality and uncertainties related to *in vivo* study data considered in the Integrated Strategy have similarly to be taken into account.

As stated in the ECHA document on uncertainty analysis, "*The underlying principle is that a tiered approach should be followed and that the amount of detail should be proportionate to the level of uncertainty and its potential impact on the risk characterisation."* (ECHA, 2012). This means that the assessment of uncertainty needs to be refined and uncertainties in the assessment reduced until an acceptable level is reached. If e.g. an analysis of the sources of uncertainty associated with use of novel tools or approaches indicates that generation of further data to address limitations in knowledge (e.g. on the relationship between the response of a human-derived *in vitro* test system to human cells *in vivo*) is unlikely to affect the outcome of the risk assessment this should be justified and documented.

This approach is dependent on the acceptable uncertainty being defined by the risk managers before the data are generated, and depending on the purpose of the risk assessment, as failure to do this could result in a number of failures. For example, only considering the level of acceptable uncertainty after data are generated could introduce bias, e.g. by deciding that a high level of uncertainty can be tolerated after poor quality data are generated. Another outcome of not deciding up-front the level of acceptable uncertainty could be paralysis of the decision making process, i.e. never being satisfied that the information are sufficient to enable a decision to be made.

Ideally, the uncertainties should be quantified, but can also be described qualitatively to support the decision making on a transparent basis.

7. References

Carthew, P., Clapp, C., and Gutsell, S. (2009). Exposure based waiving: the application of the toxicological threshold of concern (TTC) to inhalation exposure for aerosol ingredients in consumer products. *Food Chem. Toxicol.* 47(6):1287-1295. https://dx.doi.org/10.1016/j.fct.2009.02.024

Coecke S et al (2016) Chapter 5: Practical Aspects of Designing and Conducting Validation Studies Involving Multi-Study Trials. In:Eskes C, Whelan M (eds) *Validation of Alternative Methods for Toxicity Testing*, Springer, pp. 133-164

ECHA (2012) Guidance on information requirements and chemical safety assessment Chapter R.19: Uncertainty analysis. Reference: ECHA-12-G-25-EN. European Chemicals Agency, Helsinki https://echa.europa.eu/documents/10162/13632/information_requirements_r19_en.pdf/d5bd6c3f-3383-49df-894e-dea410ba4335

ECHA (2016) New Approach Methodologies in Regulatory Science Proceedings of a scientific workshop. European Chemicals Agency; Helsinki, 19–20 April 2016 https://echa.europa.eu/documents/10162/21838212/scientific_ws_proceedings_en.pdf/a2087434-0407-4705-9057-95d9c2c2cc57

EFSA (2010) European Food Safety Authority; Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal*; 8(6):1637. [90 pp.]. doi:10.2903/j.efsa.2010.1637. Available online: www.efsa.europa.eu.

EFSA Scientific Committee (2016) Guidance on Uncertainty in EFSA Scientific Assessment (Revised Draft for Internal Testing) https://www.efsa.europa.eu/sites/default/files/consultation/150618.pdf

Embry, M.R., Bachman, A.N., Bell, D.R., Boobis, A.R., Cohen, S.M., Dellarco, M., Dewhurst, I.C., Doerrer, N.G., Hines, R.N., Moretto, A., Pastoor, T.P., Phillips, R.D., Rowlands, J.C., Tanir, J.Y., Wolf, D.C., and Doe, J.E. (2014). Risk assessment in the 21st century: roadmap and matrix. *Crit. Rev. Toxicol*. 44 (Suppl 3): 6-16. https://dx.doi.org/10.3109/10408444.2014.931924

Groothuis, F., Heringa, M.,Nicol, B., Hermens, J., Blaauboer, B., Kramer, N. (2015) Dose metric considerations in *in vitro* assays to improve quantitative *in vitro–in vivo* dose extrapolations. *Toxicology* 332(5): 30-40. http://doi.org/10.1016/j.tox.2013.08.012

IPCS (2004) Industrial Programme on Chemical Safety; IPCS Risk Assessment Terminology. World Health Organization, Geneva http://www.who.int/ipcs/methods/harmonization/areas/ipcsterminologyparts1and2.pdf

Krewski, D., Acosta, D., Andersen, M., *et al.,*(2010). Toxicity testing in the 21st century: a vision and a strategy. *J. Toxicol. Environ. Health Part B* 13 (2–4), 51–138. http://dx.doi.org/10.1080/10937404.2010.483176.

Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., Würtzen, G. (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem. Toxicol.* 42(1):65-83. http://dx.doi.org/10.1016/j.fct.2003.08.006

Kroes, R., Renwick, A.G., Feron, V., Galli, C.L., Gibney, M., Greim, H., Guy, R.H., Lhuguenot, J.C., van de Sandt, J.J.M. (2007) Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. *Food Chem. Toxicol.* 45(12):2533-2562. http://dx.doi.org/10.1016/j.fct.2007.06.021

NAS (2007) Toxicity Testing in the Twenty-first Century: A Vision and a Strategy. *National Academies Press, Washington DC.* https://doi.org/10.17226/11970

NAS (2017) Using 21st Century SCeince to Improved Risk-Related Evaluations. *National Academies Press, Washington DC* https://doi.org/10.17226/24635

NAS (2012) Exposure Science in the 21st Century: A visition and a Strategy. *National Academies Press, Washington DC* https://doi.org/10.17226/13507

OECD (2007) Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, ENV/JM/MONO(2007)2. http://dx.doi.org/10.1787/9789264085442-en

OECD (2016a) **Draft** Guidance document on Good in vitro method practices (GIVIMP) for the development and implementation of in vitro methods for regulatory us in human safety assessment. http://www.oecd.org/env/ehs/testing/OECD_Draft_GIVIMP_in_Human_Safety_Assessment

OECD (2016b) Guidance Document on the Reporting of Defined Approaches to Be Used within Integrated Approaches to Testing and Assessment No. 255, ENV/JM/MONO(2016)28, OECD, Paris. https://one.oecd.org/document/ENV/JM/MONO(2016)28/en/pdf

OECD (2016c) Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to Be Used within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation, Series on Testing and Assessment No. 256, ENV/JM/MONO(2016)29, OECD, Paris. https://one.oecd.org/document/ENV/JM/MONO(2016)29/en/pdf

Renwick A and Lazarus N (1998) Human variability and noncancer risk assessment – an analysis of the default uncertainty factor. *Reg. Tox. Pharm*. 27:3-20 http://dx.doi.org/10.1006/rtph.1997.1195

SCCS (Scientific Committee on Consumer Safety)(2016)SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation 9th

revisionhttp://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_190.pdf
SCHER (Scientific Committee on Health and Environmental Risks), SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks), SCCS (Scientific Committee on Consumer Safety) (2013) Addressing the New Challenges for Risk Assessment http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_131.pdf

Williams, F.M., Rothe, H., Barrett, G., Chiodini, A., Whyte, J., Cronin, M.T.D., Monteiro-Riviere, N.A., Plautz, J., Roper, C., Westerhout, J., Yang, C., Guy, R.H. (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. Tox. Pharm* 76: 174–186 https://doi.org/10.1016/j.yrtph.2016.01.005

Worth, A., Cronin, M., Enoch, S., Fioravanzo, E., Fuart-Gatnik, M., Pavan, M., Yang, C. (2012). Applicability of the Threshold of Toxicological Concern (TTC) approach to cosmetics – preliminary analysis. COSMOS project. http://dx.doi.org/10.2788/5059