Toxicology 436 (2020) 152421

Contents lists available at ScienceDirect

Toxicology

journal homepage: www.elsevier.com/locate/toxicol

Review

The way forward for assessing the human health safety of cosmetics in the EU - Workshop proceedings

Vera Rogiers[a,*], Emilio Benfenati[b], Ulrike Bernauer[c], Laurent Bodin[d], Paul Carmichael[e], Qasim Chaudhry[f], Pieter Jan Coenraads[g], Mark T.D. Cronin[h], Matthew Dent[i], Maria Dusinska[j], Corie Ellison[k], Janine Ezendam[l], Eric Gaffet[m], Corrado Lodovico Galli[n], Carsten Goebel[o], Berit Granum[p], Heli Miriam Hollnagel[q], Petra S. Kern[r], Kirstin Kosemund-Meynen[s], Gladys Ouédraogo[t], Eirini Panteri[u], Christophe Rousselle[v], Maciej Stepnik[w], Tamara Vanhaecke[x], Natalie von Goetz[y], Andrew Worth[z]

[a] *Vice-chair of SCCS, Vrije Universiteit Brussel, Brussels, Belgium*
[b] *Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milano, Italy*
[c] *Member of SCCS, German Federal Institute for Risk Assessment (BfR), Berlin, Germany*
[d] *Member of SCCS, Alternative Energies and Atomic Energy Commission (CEA), Saclay, France*
[e] *Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedfordshire MK44, United Kingdom*
[f] *Chair of SCCS, University of Chester, Chester, United Kingdom*
[g] *Vice-chair of SCCS, University Medical Center Groningen, Groningen, the Netherlands*
[h] *Liverpool John Moores University, School of Pharmacy and Biomolecular Sciences, Liverpool, United Kingdom*
[i] *Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, MK44 Bedfordshire*
[j] *Member of SCCS, Norwegian Institute for Air Research (NILU), Kjeller, Norway*
[k] *The Procter & Gamble Company, Cincinnati, Ohio, USA*
[l] *Member of SCCS, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands*
[m] *Member of SCCS, Institute Jean Lamour (UMR 7198 CNRS Université de Lorraine), Nancy, France*
[n] *Member of SCCS, Faculty of Pharmaceutical Sciences University of Milan, Milano, Italy*
[o] *Coty, Dept of Toxicology, Darmstadt, Germany*
[p] *Member of SCCS, Norwegian Institute of Public Health, Oslo, Norway*
[q] *Dow Europe, Toxicology and Environmental Research & Consulting, Horgen, Switzerland*
[r] *Procter & Gamble Services NV/SA, Strombeek-Bever, Belgium*
[s] *Coty, Dept of Toxicology, Darmstadt, Germany*
[t] *l'Oréal, Research and Development, Paris, France*
[u] *Member of SCCS, National and Kapodistrian University of Athens, School of Pharmacy, Division of Pharmaceutical Chemistry, Athens, Greece*
[v] *Member of SCCS, French Agency for Food, Environmental and Occupational Health & Safety (ANSES), Maisons-Alfort, France*
[w] *Member of SCCS, Nofer Institute of Occupational Medicine, Łódź, Poland*
[x] *Member of SCCS, Vrije Universiteit Brussel, Brussels, Belgium*
[y] *External expert SCCS, Federal Office of Public Health, Bern, Switzerland*
[z] *European Commission, Joint Research Centre (JRC), Ispra, Italy*

Abbreviations: ADME, Absorption, Metabolism, Distribution and Excretion; ANSA, Agencies Network on Scientific Advice; AOP, Adverse Outcome Pathway; AR, Androgen Receptor; AUC, Area Under the Curve; CC, Cramer Class; Cmax, Maximal blood Concentration; CMR, Carcinogen, Mutagen, Reproductive toxic; Css, Blood concentration at steady-state; DB, Data Base; EAR, Exposure: Activity Ratio; ER, Estrogenic Receptor; ECHA, European Chemicals Agency; EFSA, European Food Safety Authority; EMA, European Medicines Agency; EPA, Environmental Protection Agency; GI, Gastro-Intestinal; HBA, Hydroxy Benzoic Acid; hERG, human Ether-à-go-go Related Gene; IATA, Integrated Approaches to Testing and Assessment; IC, Inhibitory Concentration; ICCR, International Cooperation on Cosmetics Regulation; IRIS, Integrated Risk Information System; IVIVE, Vitro to *in vivo* Extrapolation; JRC, Joint Research Centre; LOAEL, Lowest Observed Adverse Effect Level; LRSS, Long Range Science Strategy; MIE, Molecular Initiating Event; MoA, Mode of Action; MoS, Margin of Safety; MOIE, Margin Of Internal Exposure; NAM, New Approach Methodology; NGRA, Next Generation Risk Assessment; NOAEL, No Observed Adverse Effect Level; NOEL, No Observed Effect Level; NOTEL, No Observed Transcriptional Effect Level; OECD, Organisation for Economic Cooperation and Development; PBK, Physiologically Based Kinetics; PBT, Persistent, Bioaccumulative, Toxic; PBPK, Physiologically Based Pharmaco Kinetics; PBTK, Physiologically Based Toxico Kinetics; PCA, Principal Component Analysis; PK, Pharmacokinetic; POD, Point Of Departure; PTD, Para Toluene Diamine; PPD, Para Phenylene Diamine; (Q)SAR, (Quantitative) Structure-Activity Relationship; RIFM, Research Institute of Fragrance Materials; 3Rs, Replacement, Reduction, Refinement (Principle of Russell and Burch); SAR, Structure Activity Relationship; SCCS, Scientific Committee on Consumer Safety; SED, Systemic Exposure Dose; SCHEER, Scientific Committee on Health, Environmental and Emerging Risks; THR, Thyroid Hormone Receptor; (i)TTC, (internal) Threshold of Toxicological Concern; UBA, German Environmental Agency; US, EPA US Environmental Protection Agency; WHO, World Health Organization; WoE, Weight of Evidence

* Corresponding author at: SCCS - Scientific Committee on Consumer Safety, European Commission 11, rue E. Ruppert, L-2920, Luxembourg.
E-mail address: SANTE-C2-SCCS@ec.europa.eu (V. Rogiers).

https://doi.org/10.1016/j.tox.2020.152421
Received 31 January 2020; Received in revised form 21 February 2020; Accepted 25 February 2020
Available online 28 February 2020
0300-483X/

ARTICLE INFO

Keywords:
Cosmetics
New approach methodology
Animal-Free method
Replacement method
Alternative method
Next generation risk assessment
Safety assessment

ABSTRACT

Although the need for non-animal alternatives has been well recognised for the human health hazard assessment of chemicals in general, it has become especially pressing for cosmetic ingredients due to the full implementation of testing and marketing bans on animal testing under the European Cosmetics Regulation. This means that for the safety assessment of cosmetics, the necessary safety data for both the ingredients and the finished product can be drawn from validated (or scientifically-valid), so-called "Replacement methods". In view of the challenges for safety assessment without recourse to animal test data, the Methodology Working Group of the Scientific Committee on Consumer Safety organised a workshop in February 2019 to discuss the key issues in regard to the use of animal-free alternative methods for the safety evaluation of cosmetic ingredients. This perspective article summarises the outcomes of this workshop and reflects on the state-of-the-art and possible way forward for the safety assessment of cosmetic ingredients for which no experimental animal data exist. The use and optimisation of "New Approach Methodology" that could be useful tools in the context of the "Next Generation Risk Assessment" and the strategic framework for safety assessment of cosmetics were discussed in depth.

1. Introduction

Hazard identification is, besides dose-response and exposure assessment, one of the three pillars in the safety assessment of cosmetics, which is traditionally carried out using experimental animals. In the EU, the 3Rs Principle of Russell and Burch (Replacement, Reduction, Refinement) (Russell and Burch, 1959) has been introduced into the legislation of consumer products through Directive 2010/63/EC (2020) a horizontal piece of legislation on the protection of animals used for scientific purposes. For cosmetic products, for different reasons (ethical, societal, scientific and economic) the 3Rs were reduced to only 1R, namely Replacement. Testing and marketing bans were taken up into the cosmetics legislation (Regulation (EC) No 1223/2009, 2020) for the finished products and their ingredients. Consequently, the availability of animal-free methodologies became especially pressing for the safety assessment of cosmetics. Indeed, safety data for new cosmetic ingredients can be drawn from validated (or scientifically-valid) alternative methods, which from 11 March 2013 onwards are, for the purpose of cosmetics, restricted to animal-free methodology, also named New Approach Methodology (NAM) (SCCS (Scientific Committee on Consumer Safety), 2018). In line with these measures, the SCCS takes into account for the safety assessment of Annex substances (hair dyes / colorants, preservatives and UV-filters) all available toxicological information, consisting of *in chemico* data (physicochemical data), *in silico* (computational) models including (quantitative) structure activity relationship {(Q)SAR} and read-across, grouping and Physiologically Based Pharmaco Kinetics (PBPK) and Toxico Kinetics (PBTK) modelling (SCCS (Scientific Committee on Consumer Safety), 2016). These, however, need to be combined in a Weight of Evidence (WoE) approach with results obtained *via in vitro/ex vivo* testing and any existing animal data (obtained before the legislative deadlines) and, whenever available, with information generated in human studies through clinical trials and human biomonitoring. Regulatory acceptance of several non-animal approaches has been achieved for most of the lower-tier information requirements of local toxicity and acute or short-term effects, but not for systemic effects that generally become evident after long-term exposure. No validated replacement alternatives are currently available for systemic toxicity (sub-acute, sub-chronic and chronic toxicity), carcinogenicity, reproductive toxicity and the major part of toxicokinetics (EU-ANSA, 2018; Joint Research Centre, 2018). Thus, today only two main routes exist to develop a safety dossier for a cosmetic ingredient:

- In the case of a new ingredient being developed specifically for use in a cosmetic product alone, testing needs to be in compliance with the restrictions on animal testing as provided in Regulation (EC) No 1223/2009 and safety data need to be derived from NAMs.
- Where an ingredient has already existing toxicological data derived from animal tests (*e.g.* a cosmetic ingredient already on the market) that have been carried out before the regulatory testing and marketing ban became binding, these data can still be used.
- Animal test data relating to chemical substances developed for uses other than cosmetics (*e.g.* food, medicines, biocides, *etc.*) may legally be used for supporting safety assessment of an ingredient intended to be used in a cosmetic product. Also, the toxicological data obtained for guaranteeing the safety of workers may be used (ECHA (European Chemicals Agency), 2014). It should, however, be noted that there are only few examples where companies have chosen to include such post-ban animal test data.

After the deadline of 2013, no new compounds for exclusive use in cosmetics have been introduced. This clearly shows the pressing need for additional methodologies and out of the box thinking for the safety evaluation of cosmetic ingredients to avoid a roadblock to the development of new cosmetic ingredients. In this context, the recently proposed "Next Generation Risk Assessment (NGRA)" concept could be of interest for the safety evaluation of cosmetic ingredients (Berggren et al., 2017; Dent et al., 2018) (See Fig. 1). To further explore this concept, its tools and future perspectives, the SCCS Workshop on the way forward for assessing the human health safety of cosmetics in the EU, held on February 27, 2019, was focused on NGRA.

Skin sensitisation was not covered in this workshop, although for this endpoint several NAMs are available and an NGRA strategy for skin sensitisation is currently proposed by Cosmetics Europe (manuscript in preparation).

2. What is a "next generation risk assessment" and how can it be applied to cosmetic ingredients?

The 2007 report from the US National Research Council entitled 'Toxicity Testing in the 21 st Century: A Vision and a Strategy' recommended a shift away from animal studies and towards studies performed *in vitro* using human-relevant cells or tissues (National Research Council, 2007). This report argued that instead of basing a safety assessment on observed pathologies in animals, it should be based on an understanding of the concentrations that cause changes in normal cellular signalling pathways that lead to adverse effects. This report provided a vision that appeared not only desirable but given advances in molecular techniques, bioinformatics and systems biology, also achievable. The use of these techniques to modernise the science of toxicological safety evaluation has been referred to as NGRA (USEPA, 2014), and high throughput and computational modelling approaches are now central to the US EPA's strategy for improving the evaluation of chemical safety (Thomas et al., 2019). In the context of cosmetic safety evaluation, NGRA is defined as an exposure-led, hypothesis-driven risk assessment approach that integrates *in silico*, *in chemico* and *in vitro* approaches to deliver safety decisions relevant to human health without the use of animal data. NGRA shares many similarities with traditional

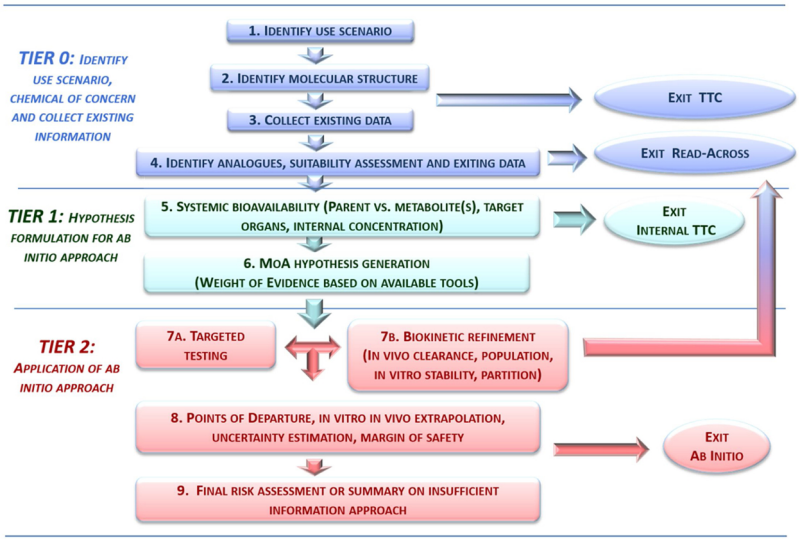

Fig. 1. Framework of the New Generation Risk Assessment (NGRA) (adopted from Berggren et al., 2017 and Dent ; et al., 2018;). TTC: Threshold of Toxicological Concern; MoA: Mode of Action.
Copyright from Elsevier, first published in Computational Toxicology, 4, 2017.

safety assessments using animal data. In both approaches, a dose or concentration that does not result in any biological activity is compared with an estimate of consumer exposure. The difference is that where animal data are used, the markers of bioactivity are generally apical endpoints or pathologies seen in animals, whilst in an NGRA the measure of bioactivity comprises a suite of *in vitro* assays. In an NGRA the *in vitro* bioactivity is compared with levels of systemic exposure in consumers (*e.g.* derived from Physiologically Based Kinetics (PBK) modelling) rather than an externally-applied dose which is used in traditional safety assessment. In either approach, a wide margin of exposure between the concentration/dose resulting in bioactivity and consumer exposures indicates that systemic bioactivity in consumers, and therefore adverse effects, are unlikely.

The principles underpinning the application of NGRA to cosmetics have been defined by the International Cooperation on Cosmetics Regulation (ICCR) (Dent et al., 2018). The ICCR determined the overall goals of NGRA as human-relevant, exposure-led, hypothesis-driven and designed to prevent harm. They also concluded that NGRA should be conducted using a tiered and iterative approach, following an appropriate literature search and evaluation of the available data, and using robust and relevant methods and strategies. Given the novelty of NGRA and the current lack of regulatory guidance on the use of many NAMs in decision-making, it is also important that the assessment should be transparently documented, explicit about the logic of the approach and sources of uncertainty.

The exposure-led nature of the safety assessment means that approaches such as the Threshold of Toxicological Concern (Yang et al., 2017) are considered at an early stage of the assessment, along with other established methods such as read-across. The SEURAT-1 workflow described in Fig. 1 (Framework of the NGRA adopted from Berggren et al., 2017 and Dent et al., 2019) fulfils many of the ICCR principles of NGRA. This workflow is exposure-led, requires all the available data to be collected and evaluated, and is tiered and iterative. Where either existing data, exposure-based waiving or read-across cannot be used to make a safety decision, more specific data need to be generated.

One major difference between traditional safety assessment and NGRA is how metabolism of the test item (either activation or inactivation) is considered. Traditional safety assessment relies on an integrated biological system (the whole animal) which takes into account the Absorption, Distribution, Metabolism, and Excretion (ADME) of the test item (usually following oral administration). A key assumption is that the processes governing the ADME of the test item are sufficiently conserved between the test species and humans and across exposure routes to make a comparison appropriate. In NGRA, these processes are broken down to ensure that the right test substances are tested *in vitro,* and the correct metrics (*e.g.* Maximal blood Concentration (Cmax), Area Under the Curve (AUC), blood Concentration at steady state (Css)) are compared in the final safety assessment. The major route of exposure for most cosmetics is dermal, and established *in vitro* assays are available to assess the kinetics of dermal absorption in relevant vehicles (Davies et al., 2010). This information, alongside knowledge of the physicochemical properties of the test item, its clearance in an *in vitro* hepatocyte model and measurements of plasma protein binding can be used to develop PBK models describing its ADME properties. Similarly, both *in silico* and *in vitro* tools can be used to predict the metabolism of the test item so that the safety assessment can be targeted on the proper entity {parent *vs.* metabolite(s)} that is linked to the Mode of Action (MoA).

In either traditional safety assessment or NGRA it is critical that the biological coverage represented by the generated data is appropriate. In other words, there needs to be enough confidence that the majority of perturbations that could lead to adverse effects have been studied before a safety decision is made. For NGRA this means including assays providing broad enough biological coverage to ensure that critical activities are not missed. Transcriptomics has a key role to play in providing sufficiently this broad biological coverage, thus giving assurance that relevant exposures will not cause systemic bioactivity that may lead to adverse effects (Thomas et al., 2011). The initial safety assessment can therefore be based on the definition of a No Observed Transcriptional Effect Level (NOTEL) (Lobenhofer et al., 2004). If the NOTEL is not reached at relevant systemic exposure levels, this provides a key conservative input for the safety assessment. Alongside the breadth of coverage provided by transcriptomics data, specific molecular initiating events (Allen et al., 2014) or signalling pathways that are known to be associated with particular adverse events (such as nuclear receptor agonism or antagonism or effects on ion channels such as hERG, human Ether-à-go-go Related Gene) should be evaluated using relevant assays. Where the NOTEL is reached at relevant exposures or where there is an insufficient margin of exposure between the bioactivity data and exposure estimate, it is important to understand the biological pathways that may be perturbed in order to conduct a higher-tier MoA driven safety assessment. This is where the ICCR principle of differentiating between biological activity and adversity ('the

assessment is designed to prevent harm') needs to be applied. However, because the assessment is tiered and iterative this step may not be needed to make a safety decision where the conservative assessment based on the bioactivity data provides a wide Margin of Safety (MoS).

A critical question for NGRA is 'what MoS is necessary or appropriate to assure safety?' Historically a margin of exposure of 100 was considered adequate to cover uncertainties relating to exposure and effects when assessing risks based on animal test data (SCCS (Scientific Committee on Consumer Safety), 2016). The figure of 100 comprises a factor of 10 for the extrapolation from test animals to an average human being (interspecies extrapolation) and another factor of 10 to account for variability within the human population (intraspecies extrapolation). These factors of 10 are further subdivided to explicitly account for inter and intraspecies differences in toxicokinetics and toxicodynamics. Therefore, one approach could be to re-define the acceptable MoS to ensure they are relevant for use with *in vitro* data and computational models instead of *in vivo* data. Indeed, the more thorough characterisation of uncertainty and variability that is possible using information-rich datasets and approaches may make some uncertainty/assessment factors redundant. For example, computer simulations of variability in systemic consumer exposure across a population has the very real potential to replace the need for uncertainty factors for toxicokinetic differences. Another approach to put MoS or exposure/activity ratios into context is to compare them with those obtained from substances with a known history of safe use. 'Dietary comparator ratios' are often used to benchmark results of *in vitro* tests against the results of animal studies, and aim to ensure that substance exposures are kept at an exposure/activity ratio that represents a region of safety (Dent et al., 2019). Using this approach, it is not necessary to predict the types of adverse effects that could occur at high doses (the 'apical endpoints' of traditional animal tests). This strategy of 'protection not prediction' is fundamental to NGRA for systemic effects, and used alongside the ICCR principles, provides the opportunity to use NAMs for safety decision making. Several case studies using the principles and approaches described above have been developed to investigate the feasibility of conducting NGRA for systemic effects of cosmetic ingredients. In one example, NGRA was performed for the hypothetical use of coumarin in face cream and body lotion at 0.1 % (Baltazar et al., 2020). In that case study, any existing *in vivo* data were discounted, and a safety assessment was performed from first principles using only *in silico* predictions and *in vitro* data. Exposure predictions were made using a tiered approach to PBK modelling using *in silico* and *in vitro* inputs (Moxon et al., 2020), informed by consumer habits and practices data specific to the population using the hypothetical products. *In vitro* Points Of Departure (PODs) were obtained from a suite of tools covering DNA damage (Toxys ToxTracker®), specific Molecular Initiating Events (MIEs) (Eurofins SafetyScreen44™), immune modulation and inflammation (BioMap® Diversity 8 Panel), and high throughput transcriptomics assays in both 2-dimensional and 3-dimensional cell cultures (BioSpyder TempO-Seq®). In addition, data from a tailored panel of biomarkers to identify effects on several stress response pathways and on mitochondrial function were used to assess the potential for coumarin to cause adverse effects as a result of causing cell stress or impaired cellular respiration (Hatherell et al., 2020). That case study highlighted how critical it is to ensure sufficient biological coverage in terms of the MoAs and cell types used, and also showed the importance of a clear tiered workflow. More publicly available examples of this type of holistic safety assessment for cosmetic ingredients are needed to refine and build confidence across all stakeholders in the application of NGRA for decision-making.

3. Some tools for NGRA

3.1. What is the Threshold of Toxicological Concern (TTC) approach and how to apply it to cosmetic ingredients?

During the past decade, many basic and some highly sophisticated *in vitro* models were developed to depict specific molecular, cellular and tissue effects of chemicals. Highly specialised *in vitro* models are used in pharmaceutical active screening to investigate whether compounds evoke specific wanted or unwanted MoAs. Most *in vitro* and *in silico* models are useful to investigate specific MoAs, but cannot provide general safety assurance in the absence of information on MoAs, including toxicokinetics and toxicodynamics, of a given compound throughout the multitude of tissues of an integrated organism. This is the point in animal data-free risk assessment where the value of read-across and TTC approaches becomes apparent. Analogue and category read-across depend on the availability of pre-2013 animal data for the read-across source.

TTC is a concept allowing, under specific conditions, to conclude without the need to run new animal studies whether a certain low level chronic human exposure is of negligible health concern or whether further work is necessary. The TTC concept includes a decision tree and series of assessments to exclude chemical structures excluded from the concept, and to address cancer and non-cancer endpoints. For information on the TTC concept, the reader is referred to some pivotal publications and reviews: Munro, 1990; Munro, 1996; Kroes et al., 2004; EPFA, 2012; EFSA/WHO, 2016. In a regulatory context, the TTC approach is being applied by agencies to flavors and fragrances, food contact material migrants, pesticide metabolites and impurities, and drug impurities. The derivation of the various exposure thresholds is based on databases of existing oral animal studies, which are used in a broad read-across within 5 classes: DNA-reactive structures, organophosphates/carbamates, and Cramer Classes I, II and III. DNA-reactive structures are assessed with a specific threshold derived from cancer study data, which are not further discussed here. The span of types of chemical structures represented in the TTC databases becomes relevant for whether TTC is an appropriate safety assessment tool or not, as read-across as a concept is mostly based on the observation that structural similarities lead to biological similarity. Hence, the objective of TTC databases is to capture toxicity data on as many chemicals as possible, to achieve that a broad range of structures can be assessed.

In the past decades, toxicological testing was prioritised for compounds which were suspected to be relevant and potent toxicants. This led to a testing and reporting bias within most toxicology databases with an over-representation of more potent compounds. Specifically, compounds traditionally used in personal care products which typically have a history of safe use and low concern for toxicity, are less abundant in toxicology databases. Similarly, a lack of representation of some structural classes had been noticed. This inspired a multi-stakeholder EU project (the COSMOS project) which set out to add cosmetic-related compounds to the TTC databases (Yang et al., 2017).

The first step was to identify which chemical structures were described by databases and the literature to be associated with personal care products/cosmetic products. This could be *e.g.* intentional ingredients, impurities thereof, contaminants or packaging material migrants. Whether an exposure to a compound is intentional or unintentional does not matter for safety assessment. Toxicology data for those cosmetics-relevant structures were then mined, curated in a database, submitted to quality control and relevance assessment, and PODs selected for threshold derivation. The data searches for the COSMOS project primarily aimed at using well established databases and data sources with primarily pre-reviewed toxicity data, for example the US EPA ToxRef DB (US Environmental Protection Agency Toxicity Reference Database) and IRIS (Integrated Risk Information System), SCCS Opinions, US FDA PAFA/CFSAN (Priority-based Assessment of Food Additives/Center for Food Safety and Applied Nutrition). Still, it was confirmed that aspects of quality control and comparability with prior existing TTC datasets need to be taken into consideration. Twenty-five percent of the PODs of the COSMOS TTC dataset comprising 552 chemicals was reviewed manually for study relevance and reliability during the step of choosing one POD per chemical. Comparison of different POD distributions demonstrated that data evaluation for the 10th percentiles and Cramer Class assignments have more impact on

Table 1
Chemical classes and 5th percentile Cramer Class PODs of selected published TTC datasetsn = number of substances /PODs in the dataset. 5th percentiles were derived from log-normal parametric distributions, except by Pinalli et al.(2011), van Ravenzwaay (2017) and Kalkhof et al.(2012), who did not report the calculation method. NOAEL = No Observed Adverse Effect Level; POD = Point of Departure; CC = Cramer Class; DB = Data Base.

Publication	Dominant chemical classes	n	5th percentile NOAELs adjusted to chronic (POD, mg/kg bw/day)		
			Cramer Class I	Cramer Class II	Cramer Class III
Yang et al., 2017: 'federated'	Cosmetic-related, packaging and pesticides	977	4.57	0.62	0.23
Munro et al., 1996	Food contact, pesticides	612	3.0	0.91	0.15
Yang et al., 2017: Munro, 2016	Some Cramer Classes corrected, harmonised assessment factors	612	4.90	1.07	0.15
Yang et al., 2017: COSMOS, 2017	Cosmetic-related & packaging	552	4.20	0.58	0.79
Tluczkiewicz et al., 2011	Industrial chemicals and pesticides	521	3.2	0.71/2.46	0.11
Kalkhof et al., 2012	German pre-REACH DB, industrial chemicals	813	2.5 (n = 69)	2.5 (n = 20)	1.3 (n = 724)
Pinalli et al., 2011	Food contact materials	232	CCI/II not reported	[£]CCIII reported in Feigenbaum et al., 2015	0.4[£] (n = 113)
Feigenbaum et al., 2015	Pesticides without carbamates and organophosphates	279	–	–	0.2
	Munro + Pinalli + pesticides with carbamates and organophosphates	840			0.15
Laufersweiler, 2012	Only reproductive and developmental endpoints. From Kroes, Bernauer, plus literature.	283	13.1 (n = 69)	1.87 (n = 11)	0.31 (n = 203)
Van Ravenzwaay, 2017	Chemicals & pesticides, developmental studies only	150§/ 537*	§rabbits: 5/9.5 maternal/dev *rats: 7.6/10 maternal/dev		

the thresholds than choice of *e.g.* time extrapolation factors or exclusion of certain structural classes. The thresholds derived for the COSMOS TTC dataset and a combined COSMOS / Munro dataset were similar to multiple other TTC non-cancer datasets, demonstrating robustness of the approach. For comparison, Table 1 lists the 5th percentiles of POD distributions for different published TTC non-cancer datasets. These PODs are either chronic study No Observable Adverse Effect Levels (NOAELs) or the duration-corrected NOAELs of subchronic or subacute studies and would be divided by an additional assessment factor of 100 to derive the respective TTC thresholds.

Chemical space comparison of the new COSMOS TTC dataset *versus* the incumbent Munro dataset (Munro, 1996) demonstrated that there was significant overlap between the datasets, but that also new chemical classes had been added to the TTC dataset, specifically cationic surfactants, hair dye components and organo-silicones.

Not surprisingly, the dataset with predominantly cosmetics-related compounds contained less potent substances and expressed higher TTC thresholds for Cramer Classes I and III. The combination of the existing Munro dataset with the COSMOS TTC dataset ('federated') resulted in a TTC dataset with a significantly higher number of structures (> 900), again increasing the probability that any new to-be-assessed structure will fall within the chemical and biological space of predictivity of the TTC thresholds. Still, it needs to be noted that several types of compounds are not amenable to assessment by TTC, as they were not included in the datasets (metals, proteins, radioactive materials) or were excluded due to extreme potency (aflatoxins, N-nitroso- and acetoxy-compounds, polyhalogenated dioxins/furans, steroids). Also, TTC thresholds refer to oral systemic toxicity, not local effects at portal of entry.

For non-volatile compounds, even if dermal exposures prevail, *in vivo* toxicity testing is typically performed by the oral route, as this route of administration results in higher systemic exposures in the majority of cases, and "in practice, oral route studies are often used for the MoS calculation to consider systemic exposure" (SCCS (Scientific Committee on Consumer Safety), 2018). The application of toxicity data obtained by oral administration to assess dermal exposure *via* personal care products/cosmetic products is the default case in personal care risk assessment and requires considerations on oral and dermal systemic bioavailability. Traditionally, the default assumption in risk assessment was that dermal absorption is lower than oral absorption so that direct comparison of dermal exposures to oral (substance-specific or TTC) threshold doses was judged conservatively {also reflected by Kroes et al. (Kroes et al., 2007)}. A more precise approach is to determine or estimate dermal and oral bioavailability for risk characterisation. This cannot only be achieved by calculations and comparison of the Systemic Exposure Dose (SED) from oral and dermal exposure {for example as laid out in the (SCCS (Scientific Committee on Consumer Safety), 2018}, but also by estimating an oral equivalent dose from the dermal exposure, which can then be compared to the oral TTC thresholds (Fig. 2).

The TTC concept is not a single tool or method, but a broad hazard characterization framework, which requires proficiency in expert judgement or (Quantitative) Structure Activity Relationship {(Q)SAR} tools for DNA-reactivity, and in Cramer Class assignment, as well as in exclusion of compounds for which TTC concept is not applicable. With that, TTC is not a layman tool, but requires broad knowledge in different aspects of hazard and risk characterisation and familiarity with the concept itself. To scientists, not involved in the area, it might seem as if there were an inconsistency between the outcomes of single projects and respective published thresholds during the last decade. However, the publication of multiple slightly different thresholds is rather an expression of progress in expanding datasets and understanding the influence of different study types, chemistries and assessment approaches. Moreover, the fact that the different thresholds were relatively comparable demonstrates that the approach is reproducible and robust.

As a general rule, TTC {and (Q)SAR} risk assessment should be based on the data which maximally combine the aspects of being up-to-date, data reliability, large database and large chemical space. Out of the datasets listed in Table 1, the federated data set of Yang et al. (Yang et al., 2017) currently has the best overlap of those criteria. The resulting TTC are for Cramer Class I (CCI), 2700 μg/person/day or 46 μg/kg bw/day, respectively; for CCII 370 μg/person/day or 6.2 μg/kg bw/day and for CCIII 140 μg/person/day or 2.3 μg/kg bw/day (body weight default value for adults is 60 kg). The reproductive-developmental datasets of Table 1 (Laufersweiler et al., 2012 and van Ravenzwaay et al., 2017) provide assurance that those specific endpoints are adequately covered when applying thresholds based on general toxicity.

As with other frameworks, the TTC concept should be constantly improved as far as possible. Regulatory applications will benefit from regular reviews by international expert groups of the progress made, to achieve consensus on if and how the established approach described by the EFSA/WHO report (EFSA (European Food Safety Authority), World Health Organization (WHO) 2016) should be adapted to progress.

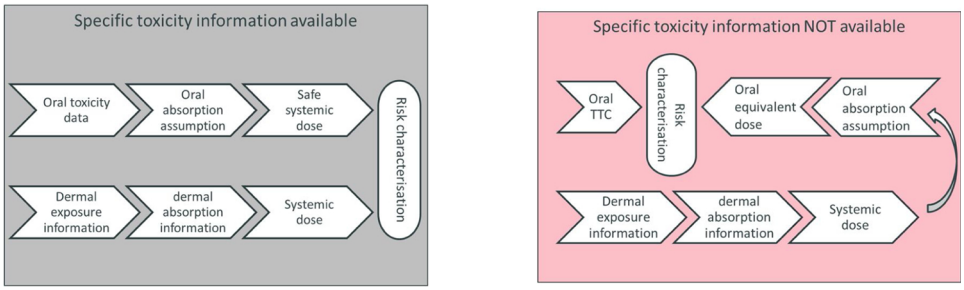

Fig. 2. A more precise approach for risk assessment by determining or estimating dermal and oral bioavailability (first pass metabolism between the oral and dermal route are not accounted for in this figure).

Within this context, it is highly advisable to maintain large, publicly accessible, curated and quality-controlled toxicology datasets – to achieve transparency in TTC and also to facilitate further method development *e.g.* in computational toxicology.

3.2. The concept of internal TTC (iTTC) and its potential use for cosmetic safety assessments

An internal Threshold of Toxicological Concern (iTTC) is *i.e.* a TTC based on plasma concentration. It is proposed as a possible evolution of the external-dose based TTC and would be a useful approach in the development of non-animal methods and as a tool to further refine the use of TTC and expand its applicability. iTTC would provide conservative exposure limits that could be utilised in exposure-based human safety assessments in the context of:

(1) refinement of de minimis exposure levels for dermal exposures;
(2) metabolism-based Structure Activity Relationship (SAR) assessments;
(3) low level aggregate exposures from different dose routes, or;
(4) *in vitro* bioactivity assays.

Several initial attempts to derive an iTTC have been made by various groups, with each taking a different approach. Partosch et al. (Partosch et al., 2015) attempted to derive an iTTC by adjusting the external NOAELs of substances from three databases by *in silico* estimates of bioavailability. The oral bioavailability prediction method used by Partosch et al. (Partosch et al., 2015) takes into account passive oral absorption but does not consider the possible impact of active transport (either influx or efflux), pre-systemic metabolism, systemic metabolism and clearance or any of the important factors that would impact internal exposure levels to target tissues after passive uptake from the gastro-intestinal (GI) tract. Animal *versus*-human differences in metabolism or other kinetic determinants for the chemicals were also not taken into account. The estimates provided were still based on external dose and not an internal exposure metric, such as plasma concentration. This approach was also utilised by Reilly et al. (Reilly et al., 2019). Chebekoue and Krishnan (Chebekoue and Krishnan, 2019) derived occupational TTCs for inhalation exposure to systemically-acting organic chemical vapors by using PBPK modelling to predict internal doses. Blackburn et al. (2019) explored the potential application of an iTTC approach to address uncertainty associated with a metabolism-based SAR assessment. The authors calculated surrogate iTTC values by utilising the 5th percentile No Observed Effect Level (NOEL) from each Cramer Class and adjusting for rat body weight and plasma volume. The authors noted that oral absorption was not considered and emphasised that the calculated values are for illustration purposes only and are not to be used in practice.

Cosmetics Europe has initiated a project to derive iTTCs, which will be based on internal exposure and will be broadly applicable to a diverse chemical space (Ellison et al., 2019). The remainder of this section will therefore highlight some of the key outcomes from this ongoing work. The proposed approach from Ellison et al. (Ellison et al., 2019) for deriving iTTCs is shown in Fig. 3.

The approach starts with identifying an existing TTC dataset which contains chemical specific NOAELs expressed as an external dose in mg/kg/day. For each chemical within the dataset, a literature search is completed to identify existing metabolism and Pharmacokinetic (PK) data and *in silico* estimates of ADME parameters are predicted. For selected chemicals, new *in vitro* ADME data (*e.g.* metabolism, permeability, binding) may be generated. Chemical-specific PBPK modelling is then conducted to convert the chemical-specific NOAEL from the toxicity study to an internal blood concentration Css, AUC, or Cmax for each chemical. The PBPK modelling uses the appropriate species, dose, and route from the toxicity study. The distribution of chemical-specific Css values will then be evaluated and an appropriately low (*e.g.* 5th percentile) Css threshold will be identified for the group of chemicals and by Cramer Class. Uncertainty factors will also be applied to the Css threshold values to derive the iTTC values that can be applied to a human safety assessment. However, additional discussion will be needed to determine which values would be appropriate, given that a comparison of an animal-derived iTTC value to a modelled human blood level for a specific exposure may help quantify extrapolation factors or uncertainty factors to account for interspecies toxicokinetics differences The chemical databases being utilised in the iTTC project are the Munro et al. (Munro et al., 1996) non-cancer TTC database, the COSMOS non-cancer TTC database (Yang et al., 2017) and the Research Institute of Fragrance Materials (RIFM) database. The Munro et al. non-cancer TTC database (Munro et al., 1996) is one of the more widely known and cited TTC databases and consists of 613 chemicals, representing a range of industrial chemicals, pharmaceuticals, food substances and environmental, agricultural and consumer chemicals likely to be encountered in commerce. The database contains data from multiple species (rat, mouse, rabbit and hamster) and all data are derived from oral exposure studies. The COSMOS database is a non-cancer TTC dataset enhanced with cosmetic and packaging materials. The toxicity studies used from the RIFM database represent flavour and fragrance materials and contain a large number of chemicals in Cramer Class II, which is a Cramer Class that has been historically under-represented in TTC databases.

As described in Ellison et al. (Ellison et al., 2019), an automated workflow was developed and utilised to locate relevant plasma concentration *versus* time data and metabolism data for the 1251 chemicals in the combined database of Munro et al. (Munro et al., 1996), COSMOS and the RIFM database). The search was performed with respect to a specific species for which the NOAEL data were collected. Manual review of the literature and extraction of the relevant data were then completed. At the conclusion of the literature search, 80 % of the chemicals had no PK or metabolism data, 10 % had PK data, 5% had *in vitro* metabolism data and 5% had both PK and *in vitro* metabolism data. The results from the literature search will be used as input into the PBPK models and help identify which chemicals need further *in vitro* metabolism testing.

Ellison et al. (Ellison et al., 2019) utilised two statistical approaches, k-means clustering and Principal Component Analysis (PCA), to map out the chemical space for the iTTC chemicals. The approaches utilised

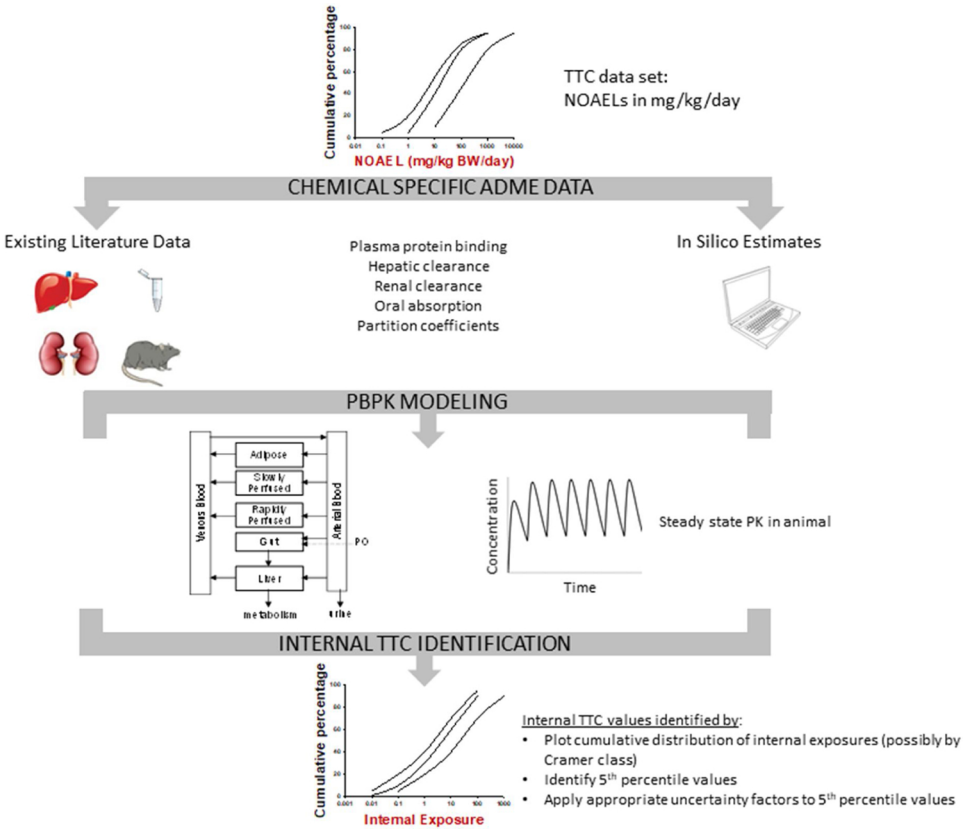

Fig. 3. Approach for deriving internal Threshold of Toxicological Concern (iTTC) values (Ellison et al., 2019). TTC: Threshold of Toxicological Concern; NOAEL: No Observable Adverse Effect Level; PK: Pharmacokinetics; PBPK: Physiologically Based Pharmacokinetics model; ADME: Absorption, Distribution, Metabolism and Excretion.
Copyright from Elsevier, first published in Regulatory Toxicology and Pharmacology, 103, 2019.

structural and molecular descriptors in addition to ADME properties as input parameters. The output from the k-means clustering and PCA analyses will help separate and group the chemicals into clusters so that representative chemicals can be selected, thus reducing the number of chemicals included in the PBPK modelling portion of the iTTC project to a more manageable number. Approximately 300 chemicals are currently estimated to be included in the final iTTC chemical database. For the chemicals included in the database, approximately 45 % will have existing PK data in the same species as the toxicity study and all will have *in vitro* metabolism. The chemicals will span across the existing TTC chemical space, will cover all three Cramer Classes, represent a broad physicochemical space and have NOAELs that span several orders of magnitude.

Ongoing work is occurring in several areas of the project. *In vitro* Caco-2 data are being collected for the iTTC chemicals and will be used as input into the PBPK modelling simulations. In partnership with RIFM and the Long-Range Research Initiative of the American Chemistry Council, cryopreserved hepatocytes (primarily rat, but also mouse, rabbit and dog) are being used to determine the *in vitro* metabolism rate for iTTC chemicals which do not have existing data in the literature. Evaluation of the PBPK modelling workflow is being conducted by comparing plasma *versus* time concentration profiles from the PBPK model to existing PK data from preclinical studies. Also, the PBPK models are being used to estimate the internal exposure associated with the NOAELs from the toxicity studies.

Upon successful completion of the project, a set of robust iTTCs will be available for use in a human safety assessment. The availability of iTTCs will expand the possible use of the TTC concept for the human safety assessment. Furthermore, due to the type of data that is needed to do this work, an added benefit of the project is that it will: provide compiled databases of existing PK and metabolism data; generate new *in vitro* Caco-2 permeability and hepatocyte metabolism data for hundreds of compounds under standardised conditions; develop PBPK modelling workflows for bottom-up *In Vitro* to *In Vivo* Extrapolation (IVIVE); characterise the magnitude of uncertainty when doing bottom-up PBPK modelling for a broad set of chemicals.

3.3. The use of Read-Across in cosmetic safety assessments - a case study example

Read-across is, possibly, the simplest of the *in silico* approaches to fill toxicological data gaps for cosmetic ingredients and materials. It is founded around the premise that similar compounds will have similar properties and activities; hence the toxicity of one molecule can be interpolated from another (Cronin et al., 2013). In terms of the making decisions about the safety of cosmetics ingredients and materials read-across approaches can be implemented at a number of places in a tiered approach to safety assessment (*e.g.* that of Berggren et al., 2017; OECD, 2017). One of the keys to implementing read-across in the absence of other data is to understand when there is sufficient evidence and justification to make a decision – if similarity can be assured and read-across is from high quality data, then a data gap may be filled early in the safety assessment process.

The structural basis of chemical similarity is at the heart of the read-across paradigm and is often simple to define. For instance, Schultz et al. (Schultz et al., 2017) demonstrated a simple approach to the grouping of n-alkanols allowing the possible read-across of PODs to a broader set of compounds. Such compounds form a natural grouping based around a common functional group with varying carbon chain lengths. The commonality of structure alone, in this case a hydroxyl group at the end of an unbranched saturated alkyl chain, does not in itself necessarily allow for read-across. Justification of the structural similarity may need to be reinforced by a number of other pieces of evidence; these are based around the following:

- Chemical structure
- Physicochemical properties and descriptors
- NAM data, including *In silico* toxicity prediction and screening

- ADME
- Existing “traditional” toxicity data and PODs

As increasing amounts of evidence are compiled there should be a commensurate decrease in uncertainty and hence increase in confidence in the read-across argument. This requires the user to decide when the data and knowledge are sufficient to justify the read-across hypothesis. *In silico* toxicology can call upon a range of resources to support information gathering. These range from the use of databases of existing toxicity data and data for NAMs, (Pawar et al., 2019), the calculation of physicochemical properties and descriptors for the molecules, the use of *in silico* toxicity prediction and screening softwares, as well as the assessment of ingredient use and overall exposure. With regard to the *n*-alkanols, the case study from Schultz et al. (Schultz et al., 2017) indicated that the 90-day repeat dose toxicity POD for 1-pentanol, being 1000 mg/kg bw per day, could be read-across to the other data for *n*-alkanols. The hypothesis for the read-across is based upon the structural similarity as well as the similar and reversible MoA of the compounds. Within the category other similarities are in terms of biokinetics, the rapid absorption from the gut and a common two-step oxidative metabolism known to occur in the liver. Within the group of *n*-alkanols, chain length clearly will affect physicochemical properties but in this study it is assumed not to affect or influence mammalian acute and subchronic toxicity *via* oral exposure.

In order to evaluate the robustness of the read-across for n-alkanols, as well as for other read-across cases, Schultz and Cronin (Schultz and Cronin, 2017) summarised the limitations of read-across. Some of the key areas where further information was required included the justification of similarity and read-across hypothesis, a better understanding of the quality of underlying data and more information on toxicokinetics. For many read-across scenarios, further information can be obtained from knowledge of mechanisms of toxic action and/or AOPs. In addition to the information on chemical structure, *in silico* screening and data from NAMs, such as ToxCast and Tox21 can support the mechanistic hypothesis or relevant AOP.

In order to understand better the uncertainty in read-across for toxicological data gap filling, Schultz et al. (2019) have defined twelve areas of uncertainty were and a series of questions were illustrated through examples. The main categories of uncertainty defined as: (1) the regulatory use of the prediction, which dictates the acceptable level of uncertainties, (2) the data for the apical endpoint being assessed (these must be of sufficient quality and relevance for data gap filling), (3) the read-across argumentation (including mechanistic plausibility, completeness of the supporting evidence, robustness of the supporting data, and WoE approach) and (4) the similarity justification. This allows areas of uncertainty to be highlighted and hence more information can be included to decrease uncertainty and increase confidence (Schultz et al., 2019). There is a particular focus on aspects affecting toxicodynamics and toxicokinetics. This implies scenarios where there are multiple streams of evidence of varying quantity and quality across a group of compounds. New methods to assess these streams of data are being developed including Dempster-Shafer Theory which provides a more quantitative estimate of uncertainty which could be applied for risk assessments (Rathman et al., 2018). The key question to answer at this point is what is the acceptable level of uncertainty, and how this can be defined, to make a specific decision.

It is clear that more work and effort is required in the area of read-across to fill toxicological data gaps. There is a need to raise awareness and audit the types of computational methods and databases that may be useful in the safety assessment of cosmetic ingredients. More work is required in the identification and quantification of uncertainty related to read-across and how these have an impact on the confidence we can place in a decision. In addition, greater emphasis will be placed on the methods used to integrate and combine data from different structural analogues when decisions must be made based on their chemistry and toxicology.

3.4. Building the weight of evidence through in silico methods - integrating read-across and (Q)SAR models for safety assessment of cosmetic ingredients

In silico models are rapidly moving from a set of separate tools, to a complex architecture, where different software modules are integrated. This follows the effort to be close to the regulatory and industrial requirements, which refers to the chemical safety. Recently European Food Safety Authority (EFSA) published a guidance document (EFSA (European Food Safety Authority), 2017) which provides a general framework to integrate data of heterogeneous nature, including results from *in silico* models and read-across. Based on this guidance, a review addressed the different ways to merge values from *in silico* models and read-across (Benfenati et al., 2019). Indeed, the interest of the stakeholders is not on the methodology *per se*, but on the substance evaluation. From this perspective, any piece of information should be exploited. It is very beneficial to integrate the data from *in silico* models and read-across tools: the presence of similar compounds with the property value close to what is predicted by the *in silico* model surely increases the confidence on the final assessment. Read-across provides the experimental evidence related to a very local toxicological and structural area of the chemical space, while the *in silico* model gives a more general view of the toxicological processes, codified within the model. Software programs which integrate *in silico* models and read-across have been recently developed; the VEGAHUB platform is an example of a freely available (www.vegahub.eu) platform with tools integrating *in silico* models and read-across.

Furthermore, the integration among multiple tools is now pointing towards the development of frameworks wrapping models for exposure and hazard within the same system. This is the target of the EU project VERMEER (www.life-vermeer.eu). Within this project, a single software system has been designed, which predicts the skin permeation using *in silico* models, and then predicts whether the substance is a CMR (Carcinogen, Mutagen, Reproductive toxic), and skin sensitiser. The Cramer classes and NOAEL and Lowest Observed Adverse Effect Level (LOAEL) values are also predicted. Based on the ingredient concentration (provided by the user), and on the results of these exposure and hazard models, the safety of the ingredients at a certain concentration is assessed. This software will be freely available within VEGAHUB (where anyhow several separate models to be wrapped are already available). It is noticed that the information on the permeation (which can be skipped by the user) moves towards a risk assessment perspective related to the internal dose. This shift requires more studies. Along this direction, a recent project, funded by EFSA, called Optitox, is developing toxicokinetic models to be combined with the hazard *in silico* models. The expectation is to have more accurate predicted values.

An interesting advantage of the *in silico* models is the possibility to run them on large collections of substances. In a recent study, about 20,000 cosmetic ingredients of botanical origin have been screened for mutagenicity; multiple *in silico* models have been used, and their results integrated, providing an overall value with the associated uncertainty, related to the level of agreement among the models (Raitano et al., 2019). A higher level of integration among different models can be achieved for prioritisation purposes, combining multiple values on multiple endpoints. The JANUS software is an example. This software has been developed within a project funded by the German UBA (German Environmental Agency), and combines 48 different *in silico* models, to get a single CMR, Persistent, Bioaccumulative, Toxic (PBT) and endocrine disruption assessment. In this case, for carcinogenicity and reproductive toxicity, quantitative assessment is provided, offering the way to sort the substances to be prioritised in a continuous way. For this purpose, the *in silico* models for carcinogenicity include four classifiers and two regression models, based on the slope factor for inhalation and oral exposure. Further challenges for *in silico* models are the possibility to plan safer substances, substituting those with higher concern. This is under development within the VERMEER project,

exploiting software which identifies the structural alerts of concern, and replaces them with other fragments not associated with adverse effects.

3.5. Building confidence in using new approach data in an NGRA strategy

The ICCR (Dent et al., 2018) proposed 9 principles underpinning the use of NGRA in the risk assessment of cosmetic ingredients. The overall goals should be '(1) *human-relevant*, (2) *exposure-led,* (3) *hypothesis-driven and* (4) *designed to prevent harm.*' The assessment should '(5) *follow an appropriate appraisal of all existing information and use* (6) *a tiered and iterative approach*, (7) '. '(8) *Sources of uncertainty should be characterized and documented and* (9)' (Dent et al., 2018). Among these principles, the last two are particularly relevant for building confidence in the use of new approach data. At present, however, there is no precedent, either within regulatory dossiers or the scientific literature, to illustrate how these principles should be applied in practice. Yet, there is a plethora of guidance on how to characterise and communicate uncertainties in regulatory assessments, including documents from the (EFSA (European Food Safety Authority), 2018a, b; EFSA (European Food Safety Authority), 2019), the ECHA (European Chemicals Agency) (2012), the EU Scientific Committee on Health, Environmental and Emerging Risks (SCHEER (Scientific Committee on Health, Environmental and Emerging Risks) 2018) and the World Health Organisation (World Health Organization (WHO), 2017). However, it is a challenge to navigate through all of the documents. To partially address this problem, a project within the Organisation for Economic Cooperation and Development (OECD) (2005), 2007; Organisation for Economic Cooperation and Development (OECD) (2014), 2017 is ongoing to develop an overview document on concepts and available guidance on Integrated Approaches to Testing and Assessment (IATA). Guidance addressing uncertainty assessment is covered in this overview. At the same time, mature proposals are emerging from the scientific community regarding the uncertainty assessment of different types of NAMs, including (Q)SARs (Cronin et al., 2019) and read-across (Schultz et al., 2019). Established principles and procedures for method validation (OECD, 2005) are also relevant, since the validation of an individual method, be it a conventional *in vitro* test or an advanced microphysiological device, should not be regarded as an end in itself, but rather as a source of information that is propagated further through uncertainty assessment frameworks for IATA, including NGRA approaches (Fig. 4).

The generation and integration of data from different types of NAMs is covered elsewhere in this paper. To date, NGRA case studies have focused on the integration of data, but less so on the treatment of uncertainties. How to propagate and integrate the uncertainties deriving from the use of multiple NAMs thus remains a scientific challenge. However, there are lessons to be learned from different areas of mathematical modelling, including (Q)SAR (OECD, 2007), PBK modelling (EMA (European Medicines Agency), 2018) and statistics where approaches such as Bayesian modelling are well established (Bois and Diack, 2005) and experience is being gained with new approaches such as Dempster-Shafer theory (Rathman et al., 2018).

In conclusion, the characterisation and propagation of uncertainties (building confidence) resulting from the use of NAMs warrants further attention within NGRA case studies. This experience should in turn inform the future development of regulatory guidance on how to apply NGRA to cosmetic ingredients.

3.6. Case study of the safety assessment of a new hair dye based on NAMs

In general, a pre-requisite for structural analogue-based toxicological assessment is to identify and rank potential analogues for their suitability. Computer tools and expert judgement are the basis for identifying similarities and establishing plausibility in regard to confidence and justification. Published guidelines and frameworks and a comprehensive number of published case studies outline integrative processes for ranking structural similarity, reactivity similarity, metabolism similarity and similarities in physicochemical properties (ECHA (European Chemicals Agency), 2017; OECD, 2014; Przybylak et al., 2017; Desprez et al., 2018). Confidence and acceptance in read-across prediction is achieved by an in-depth evaluation and integration of uncertainty areas including those related to read-across argumentation, justification and data quality (Blackburn and Stuard, 2014; Schultz et al., 2019).

For the present read-across case study with aromatic amine hair dyes of the *p*-phenylenediamine (PPD) type, we considered the PPD-derivative toluene-2,5-diamine (PTD) as the target chemical without 90-day repeated dose toxicity data. Subsequently, the individual steps of the structural analogue-based toxicological assessment are briefly described.

3.6.1. Selection of source chemicals by structural similarity

The selection of the source chemicals was based on three key criteria. Firstly, structural similarity of more than 50 % was assessed by using the ChemIDplus structural similarity tool (https://chem.nlm.nih.gov/chemidplus/) with the 1,4-benzenediamine (PPD) free base structure in combination with the 2-methyl-side chain (target structure) and the target structure 2-methyl-1,4-benzenediamine (PTD) alone. Secondly, retrieved compounds were manually filtered to match the benzenediamine structure with amino groups in positions 1 and 4 (see Fig. 5, red circles) and side chains in position 2 (blue circles). This was considered to set appropriate limits for the potential biological reactivity (see Table 2 below). Thirdly, compounds with relevant data for repeated dose toxicity were selected.

3.6.2. Consideration of similarities in metabolism and reactivity

Based on the structural similarity selection, we used the data-rich source compounds PPD, 2-methyl-PPD and hydroxyethyl-PPD as read-

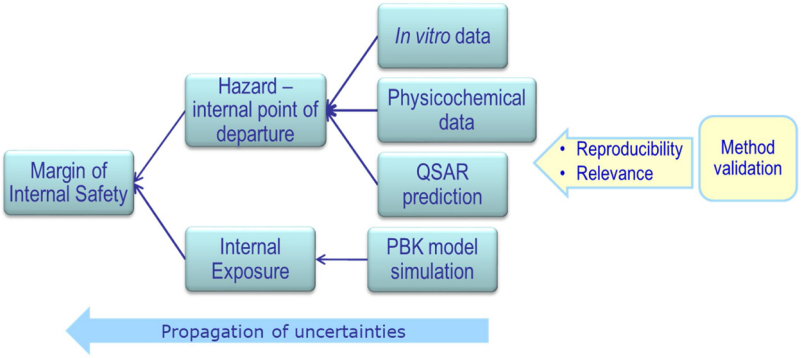

Fig. 4. Method validation within the context of Next Generation Risk Assessment. (Q)SAR = (Quantitative) Structure Activity Relationship.
The SCCS takes, besides validated methods, also scientifically valid methodology into consideration when supported by relevant and robust data.

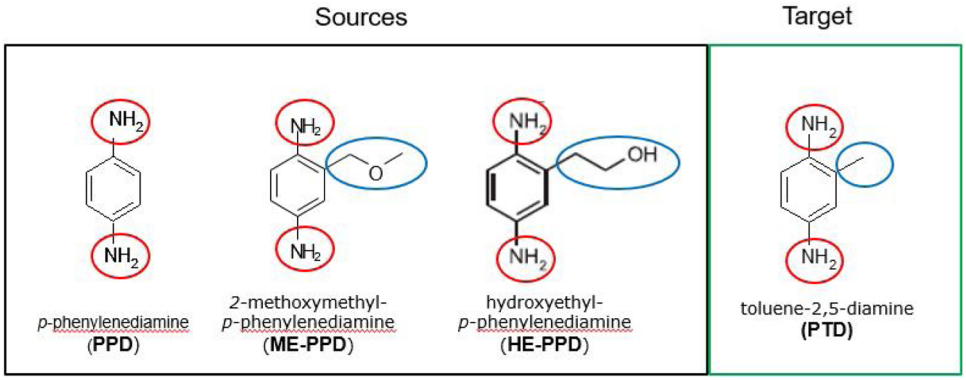

Fig. 5. Structural features of target and source compounds. Structural similarity (using ChemIDplus) was 60 % for PPD, 52 % for ME-PPD and 54 % for HE-PPD.

across analogues for the target chemical PTD. The corresponding data on repeated dose toxicity including study guidelines, dosing, toxic effects and derived NOAELs are compiled in Table 2 above. For the source compounds and the target, robust and consistent *ex vivo* and *in vitro* data are available (see Table 2) indicating that these molecules can be metabolically inactivated by *N*-acetyl-transferase-1 in human skin (first pass effect) and systemically in the liver by *N*-acetyltransferase-2, confirming the close similarity in biotransformation *via* the same metabolic pathway.

In chemico and *in vitro* experiments for biological reactivity with protein (cysteine reactivity) and DNA (Comet assay) confirm that under conditions of maximized activation cysteine and DNA reactive intermediates are generated (see Table 2). Mechanistically, this corresponds to the reactivity of Pre-Michael acceptors (Schultz et al., 2009), while the *N*-acetylated derivatives exhibit reduced biological reactivity towards cysteine and DNA (see Table 2) consistent with their detoxification through *N*-acetylation.

3.6.3. Assessment of read-across plausibility regarding confidence

For this specific assessment a high confidence categorization was applied based on the data quality. Three suitable analogues with contributing data linked to highly robust datasets for repeated dose toxicity (90-day studies), a high concordance in the observed toxicity and the resulting NOAELs have been identified (see Table 2). This high confidence level is further supported by a similar biological reactivity pattern towards cysteine (including similar depletion rate for target and analogue PPD). The read-across prediction to PTD is therefore considered equivalent to standard testing. Subsequently, PPD is selected here as the analogue with highest biological reactivity and the highest severity of the toxic effect for repeated dose toxicity. Therefore, the rat oral 90-day repeated-dose NOAEL for PPD of 8 mg/kg body weight/day is read-across to fill data gaps for PTD.

3.6.4. Considering systemic exposure

Exposure under application conditions of hair dyeing occurs on the skin (scalp) at low concentrations in the range of 2% (SCCS (Scientific Committee on Consumer Safety), 2010, 2012a; SCCS (Scientific Committee on Consumer Safety), 2012b, 2013), and detoxification by *N*-acetylation to non-reactive derivatives is the key biotransformation (see Table 2).

By assessing metabolic transformation rates *in vitro* in human skin explants (applying typical hair dye use conditions) as well in human keratinocyte cultures (HaCaT cell line) and human hepatocyte cultures, we estimated the remaining systemic exposure (AUC) to the parent compounds PPD (analogue) and PTD (target). For both compounds the estimation revealed similar AUC values. For PPD the estimated AUC data were found to be in the same order of magnitude as those published for human hair dye consumers (for details see Manwaring et al., 2015). This finding was considered to further support the selection of PPD as the most relevant analogue for read-across to PTD.

3.7. Case study of parabens as a cosmetic preservative in the EU project EU-ToxRisk based on NAMs

3.7.1. Introduction

NAMs can be used to support the safety assessment of cosmetics in different ways, depending on the question that needs to be answered. Case studies play a key role in exploring the value of NAMs in different types of safety assessment and the paraben case study is one such example in the field of systemic toxicity. Performed in collaboration between members of the Cosmetics' Europe Long Range Science Strategy (LRSS) and the Horizon 2020 EU-ToxRisk consortium (https://www.eu-toxrisk.eu/) it aims at exploring the value of NAMs in read-across for reproductive toxicity endpoints.

In a traditional systemic toxicity assessment, different endpoints are evaluated quantitatively (based on hazard identification and dose response) including repeated dose general target organ toxicity, reproductive toxicity, and developmental toxicity. A read-across assessment includes evaluation of these related systemic endpoints for the source chemical(s) in lieu of data for the target chemical(s) and traditionally uses *in vivo* data to "predict" the properties of a related compound(s) having a data gap. Historically, it was based on the use of structural features to predict hazard but that has evolved towards consideration of physicochemical characteristics, metabolism, and reactivity in determination of analogue suitability. More recently, NAMs have been shown to add value in informing read-across hypotheses and underpinning the suitability assessment of analogue(s). Efforts are underway at Cosmetics Europe to evolve read-across in this direction.

The purpose of this case study was to assess the *value added by NAMs in safety assessments based on read-across*. The following questions were therefore addressed:

- Can NAMs be used to support a read-across hypothesis for low toxicity compounds?
- Can NAMs be used to perform human-relevant risk assessment and support decision making?

To address these questions, data-rich compounds (parabens) with a clear SCCS safety assessment available were selected. Other criteria that were accounted for were: the wide use of the compounds in cosmetics *via* the dermal route, the availability of internal exposure data (to corroborate the conservatism of exposure predictions) and the availability of relevant analytical methods for characterisation. The case study was exposure led, with consideration of aggregate exposure for the short chain parabens and the hypothesis that the parent compounds have a similar MoA and are hydrolyzed to yield the same primary

Table 2
Available data for source compounds (analogues) and Target.

INCI name of chemical (abbreviation)	*p*-phenylenediamine (PPD)	2-methoxymethyl-PPD (ME-PPD)	hydroxyethyl-PPD (HE-PPD) sulfate	TARGET toluene-2,5-diamine (PTD)
Reference	8	9	10	11
CAS# free base	106-50-3 624-18-0 (dihydrochloride)	337906-36-2 337906-37-3 (sulfate)	−93841-25-9 (sulfate)	95-70-5 615-50-9 (sulfate)
MW [g/mol] free base	108.14 Dihydrochloride 181.07	152.20 Sulfate 250.28	152.20 Sulfate 250.28	122.17 Sulfate 220.25
Water solubility [g/l] Log Pow:	Free base ~10 Free base -0.31	Free base 284 Free base:-0.65	Sulfate 51.2 Sulfate 0.07	Sulfate 5 Free base 0.16; Sulfate 0.74
Repeated dose toxicity				
OECD guideline	408	408	408	
Dosing [mg/kg bw/day]	0, 2, 4, 8 and 16	0, 10, 30 and 90	0, 25, 35 and 55	
Toxicity	At 16: liver and kidney weight increase, myodegenerative effects in two animals	At 90: marginally increased activity in liver enzymes (AST), increased liver weight/ hepatocellular hypertrophy	At 55: increased activity in liver enzymes (AST, ALT)	
NOAEL [mg/kg bw/day]	8 (free base)	90 (sulfate)	35 (sulfate)	
Oral bioavailability	High (rats)	High (rats)	High (*in vitro permeability*) [Obringer et al., 2016]	High (*in vitro permeability*) [Obringer et al., 2016]
Metabolism [Manwaring et al., 2015]				
Human hepatocytes Human skin/ keratinocytes	Mono/Di-acetylation of NH_2 group(s)/no oxidation/ activation Mono/Di-acetylation/ no oxidation/ activation	Mono-acetylation of NH_2 group /no oxidation/ activation Mono-acetylation/ no oxidation/ activation	n.a. Indication for Mono/Di-acetylation of NH_2 group(s)	Mono/Di-acetylation of NH_2 group(s)/no oxidation/activation Mono/Di-acetylation/ no oxidation/activation
Biological reactivity[Goebel et al., 2014; Zeller and Pfuhler, 2014]				
Cystein reactivity at 200 μM DNA reactivity in Comet assay (% damage above control)	High (80 % depletion) High (7 fold above control)	Medium (25 % depletion). a.	n.a.. a.	High (80 % depletion) High (17 fold above control)
Effect of N-acetylation Cystein reactivity at 200 μM NA reactivty in Comet assay	Mono/Di-acetylated PPD Low (< 10 % depletion) Low (6 fold lower than parent)	Mono-acetylated ME-PPD Low (< 10 % depletion) n.a.	n.a. n.a. n.a.	Mono/Di-acetylated PTD Low (< 10 % depletion) Low (8 fold lower than parent)

metabolite; p-hydroxybenzoic acid (4-HBA). The SEURAT-1 safety assessment framework from Berggren et al. (Berggren et al., 2017; discussed in OECD, 2017) was used to run this case study as it offered the means to structure knowledge and data in a logical sequence and integrated manner, allowing for decisions to be based on exposure and MoA.

The problem addressed in this assessment was an assumed data gap for reproductive toxicity for propylparaben (available data were not taken into consideration). Data from structurally similar source chemicals, methyl-, ethyl- and butylparaben, which flank the target chemical were used in a read-across approach to fill this data gap. The decision context for the assessment was to perform a risk assessment of propylparaben, as used in different cosmetic product types.

3.7.2. Strategy used for the assessment

Read-across was used to address this data gap and to derive a POD that was subsequently used to calculate a Margin Of Internal Exposure (MOIE) building on the systemic exposure predictions and relative potency data obtained using the NAMs.

Uncertainties were assessed and an overall conclusion on the use of NAMs to support read-across and perform a human-relevant risk assessment was reached. Supporting information based on the comparison of human exposure estimates to *in vitro* activity concentrations was also explored.

The formation of a MoA hypothesis was central to this case study. The available *in vivo* data on the source substances (methyl, ethyl and butyl paraben) as well as physicochemical properties, *in silico* predictions and *in vitro* data on the source and target substances were evaluated to arrive at the hypothesis that short chain parabens with alkyl chain lengths of 1–4 carbons have a similar biological activity (low toxicity with weak estrogenic activity), and that their common major metabolite 4-HBA shows no appreciable estrogenicity. It was also hypothesised that alkyl chain length differences would result in differences in bioavailability and a predictable potency trend. *In silico* alerts were also found for Androgen Receptor (AR) and Thyroid Hormone Receptor (THR) binding that were also addressed.

In order to test the MoA hypotheses of weak estrogenic activity and the potential for interactions with AR and THR, a suite of robust and human-relevant transcriptional activation assays was performed, which provided dose-response information. Data were generated for both the target and source substances and 4-HBA in the CALUX EATS activity panel (BioDetection Systems, NL) in the presence and absence of S9. This demonstrated that metabolism of the esters to the acid resulted in a loss of the weak endocrine activity. This enabled the derivation of potency scaling factors (Toxcast ER model), which were applied in the safety assessment to a conservative *in vivo* rat NOAEL for butylparaben of 2 mg/kg/day (Fisher et al., 1999).

Alongside considerations of MoA, the exposure assessment followed a tiered approach, moving from worst-case deterministic aggregate exposure assessment based on the applied dose to a higher tier assessment that takes into account a probabilistic model of co-use of different paraben-containing products and internal exposure predictions from a PBK model. This allowed to simulate the internal exposure to both humans and rats.

The overall safety evaluation was therefore based on the integration of the PBK predictions (which included metabolism of the parent parabens to their common metabolite) and the potency ranking performed using the CALUX data.

The MOIE using the worst-case deterministic aggregate exposure value was 315, and 887 when using the probabilistic exposure modelling.

This case study can therefore be considered a 'next generation read-across' that is protective of human health and was well aligned with the SCCS opinion that concluded on the safe use of propylparaben.

3.7.3. Supporting information

As a way to further examine whether the NGRA was protective of human health, the systemic exposure estimates were compared directly to the IC50 Calux *in vitro* activity levels (following the approach described in Wetmore et al. (2012), 2015. Systemic exposure estimates were 2–5 orders of magnitude lower than the concentrations at which *in vitro* activity (estrogenic receptor agonism or androgenic receptor antagonism) was observed. Exposure Activity Ratios (EARs) were also generated by comparing the half maximal Inhibitory Concentration (IC50) CALUX *in vitro* levels to the estimated human plasma concentrations. Dietary compounds, genistein and diindolylmethane, which are known for their history of safe use, were then used as comparative benchmarks (Becker et al., 2015; Dent et al., 2019). Anti-androgenic and estrogenic activity from propyl paraben cosmetic use was lower than that coming from dietary diindolylmethane and genistein, with differences observed of 1–3 orders of magnitude, respectively. Further, we also evaluated the transcriptional response of MCF7 cells to the exposure to each of the parabens, which showed that each of the parabens shares a high degree of similarities across the category members.

This further increases confidence that current cosmetic uses are protective of human health.

4. Conclusions of the parabens case study

The evaluation of the NAM data in this case supports the safety assessment regarding the reproductive toxicity potential of propylparaben from dermal exposures.

- The OECD published workflow that foresaw such use of NAMs in NGRA was utilised and illustrated.
- NAMs (*in silico* alerts, Toxcast data, nuclear gene reporter assays, transcriptomics data, skin and liver metabolism, PBPK modelling) added value in this case study: the relative potency information and systemic exposure were useful for deriving a MOIE for the safety assessment.

Probabilistic modelling was used to further refine external exposure estimates and generate more realistic internal exposure values for safety assessment. More refinements may be possible.

5. Concluding remarks

The workshop provided an up to date overview of selected NAMs and strategies for the safety assessment of cosmetic ingredients. Speakers and participants actively and openly shared their (personal) views, which allowed a good scientific exchange. For the SCCS, this workshop was timely since the safety assessment of new cosmetic ingredients will have to rely on new non-animal concepts. These may include NGRA and WoE approaches in which NAMs, concepts such as TTC and/or iTTC and read-across are combined with historical animal data. Animal data that are generated for another regulatory purpose than cosmetics may also be included in the safety assessments, but some companies choose to not use any post-ban animal test data. Progress has clearly been made, but more examples are needed to create confidence that performing NGRA as described above is protective, also for new compounds. Concern is mainly due to lack of suitable NAMs for the more complex endpoints, *e.g.* repeated dose toxicity, developmental toxicity, biokinetics and carcinogenicity. However, experience is growing and with the help of more case studies and development of practical workflows, solutions for complex endpoints will also be explored. Also, concerns were raised that methods that are available in the OECD test guidelines program are validated for a limited set of standard chemicals and not for a sufficiently large chemical space *e.g.* nanomaterials, which complicates the safety assessment of cosmetic ingredients that are used in the nanoform.

The NGRA concept may offer an interesting platform for dealing with the safety assessment of cosmetics and their ingredients. During the plenary discussion, it was recommended to share practical solutions, for example, by using case studies in practical exercises or courses. This will facilitate learning by doing and hands-on experience with new methods and approaches. It was further stressed that the development process of NAMs and their application should be monitored closely, preferably by a multidisciplinary supervisory group of experts. Close interactions between the different actors in this field is essential to get a common understanding of the novel approach. This also allows feedback from the regulatory side on their specific needs and expectations. Especially, the case studies of data-rich substances presented during the workshop were illustrative of the process followed. To further evaluate the usefulness of this approach, more detailed information on the case studies is needed, including an assessment of the underlying confidence. It was suggested to the cosmetic industry to pro-actively take up the challenge and develop a complete dossier of a new compound based on the emerging NGRA concept. This would provide insights in potential knowledge gaps and provide experience in the topic to risk assessors on both industry and regulatory sides.

The following persons contributed to the case study of parabens: Annette Bitsch (Fraunhofer), Dagmar Bury (L'Oréal), Harvey Clewell (Scitovation), Mark Cronin (Liverpool John Moores University), Tom Cull (Unilever), Matt Dent (Unilever), Bertrand Desprez (Cosmetics Europe), Alina Efremento (Scitovation), Corie Ellison (P&G), Stefania Giammanco (Crème), Eric Hack (Scitovation), Nicola *J.* Hewitt (Cosmetics Europe consultant), Gerry Kenna (Cosmetics Europe consultant), Martina Klaric (Cosmetics Europe), Reinhard Kreiling (Clariant), Dinant Kroese (TNO), Tony Long, (Lhasa), Catherine Mahony (P&G), Enrico Mombelli (Ineris), Jorge Naciff (P&G), John O'Brien (Crème), Andreas Schepky (Beiersdorf), Sarah Tozer (P&G), Bob Van de Water, (University of Leiden), Bart Van der Burg (BDS).

6. Disclaimer

This report reflects the content of the workshop and the presentations given by the different speakers. It does not necessarily reflect the opinion of the SCCS.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.tox.2020.152421.

References

Allen, T.E.H., Goodman, J.M., Gutsell, S., Russell, P., 2014. Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. Chem. Res. Toxicol. 27 (12), 2100–2112. https://doi.org/10.1021/tx500345j.

Baltazar, M.T., Cable, S., Carmichael, P.L., Cubberley, R., Cull, T.A., Delagrange, M., Dent, M.P., Kukic, P., Li, H., Li, M.Y., Middleton, A.M., Moxon, T.E., Nathanai, l A., Nicol, B., Pendlington, R., Reynolds, G., Reynolds, J., White, A., Westmoreland, C.A., 2020. Next generation risk assessment case study for coumarin in hypothetical cosmetic products. Toxicol. Sci submitted.

Becker, R.A., Friedman, K.P., Simon, T.W., Marty, M.S., Patlewicz, G., Rowlands, J.C., 2015. An exposure, activity profiling method for interpreting high-throughput screening data for estrogenic activity—proof of concept. Regul. Toxicol. Pharmacol. 71 (3), 398–408. https://doi.org/10.1016/j.yrtph.2015.01.008.

Benfenati, E., Chaudhry, Q., Gini, G., Dorne, J.L., 2019. Integrating *in silico* models and read-across methods for predicting toxicity of chemicals: a step-wise strategy. Environ. Int. 131, 105060. https://doi.org/10.1016/j.envint.2019.105060.

Berggren, E., White, A., Ouedraogo, G., Paini, A., Richarz, A.-N., Bois, F.Y., Exner, T.,

Leite, S., van Grunsven, L.A., Worth, A., Mahony, C., 2017. Ab initio chemical safety assessment: a workflow based on exposure considerations and non-animal methods. Comput. Toxicol. 4, 31–44. https://doi.org/10.1016/j.comtox.2017.10.001.
Blackburn, K., Stuard, S.B., 2014. A framework to facilitate consistent characterization of read-across uncertainty. Regul. Toxicol. Pharmacol. 68 (3), 353–362. https://doi.org/10.1016/j.yrtph.2014.01.004.
Blackburn, K.L., Ellison, C.A., Stuard, S.B., Wu, S., 2019. Dosimetry considerations for *in vivo* and *in vitro* test data and a novel surrogate iTTC approach for read-across based on metabolites. Comput. Toxicol. 10, 145–157. https://doi.org/10.1016/j.comtox.2018.08.005.
Bois, F.Y., Diack, C.T., 2005. Uncertainty analysis: the bayesian approach. In: Edler, L., Kitsos, C.P. (Eds.), Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment. John Wiley & Sons, Inc., pp. 255–266. https://doi.org/10.1002/0470857706. ch15.
Chebekoue, S.F., Krishnan, K., 2019. Derivation of internal dose-based thresholds of toxicological concern for occupational inhalation exposure to systemically acting organic chemicals. J. Occup. Environ. Hyg. 16 (4), 308–319. https://doi.org/10.1080/15459624.2019.1568445.
Cronin, M.T.D., et al., 2013. An introduction to chemical grouping, categories and read-across to predict toxicity. In: Cronin, M.T.D. (Ed.), Chemical Toxicity Prediction: Category Formation and Read-Across. The Royal Society of Chemistry, Cambridge, pp. 1–29. https://doi.org/10.1039/9781849734400-00001.
Cronin, M.T.D., Richarz, A.N., Schultz, T.W., 2019. Identification and description of the uncertainty, variability, bias and influence in quantitative structure-activity relationships (QSARs) for toxicity prediction. Regul. Toxicol. Pharmacol. 106, 90–104. https://doi.org/10.1016/j.yrtph.2019.04.007.
Davies, M., Pendlington, R., Page, L., Roper, C., Sanders, D., Bourner, C., Pease, C., MacKay, C., 2010. Determining epidermal disposition kinetics for use in an integrated nonanimal approach to skin sensitization risk assessment. Toxicol. Sci. 119 (2), 308–318. https://doi.org/10.1093/toxsci/kfq326.
Dent, M., Teixeira, A.R., Amores Da Silva, P., Ansell, J., Boisleve, F., Masato, H., Hirose, A., Kasai, Y., Kern, P., Kreiling, R., Milstein, S., Montemayor, B., Oliveira, J., Richarz, A., Taalman, R., Vaillancourt, E., Verma, R., Posada, N.V.O.C., Weiss, C., Kojima, H., 2018. Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. Comput. Toxicol. 7, 20–26. https://doi.org/10.1016/j.comtox.2018.06.001.
Dent, M., Li, H., Carmichael, P.L., Martin, F.L., 2019. Employing dietary comparators to perform risk assessments for anti-androgens without using animal data. Toxicol. Sci. 167 (2), 375–384. https://doi.org/10.1093/toxsci/kfy245.
Desprez, B., Dent, M., Keller, D., Klaric, M., Ouédraogo, G., Cubberley, R., Duplan, H., Eilstein, J., Ellison, C., Grégoire, S., Hewitt, N.J., Jacques-Jamin, C., Lange, D., Roe, A., Rothe, H., Blaauboer, B.J., Schepky, A., Mahony, C., 2018. A strategy for systemic toxicity assessment based on non-animal approaches: the Cosmetics Europe Long Range Science Strategy programme. Toxicol. Vitr. 50, 137–146. https://doi.org/10.1016/j.tiv.2018.02.017.
ECHA (European Chemicals Agency), 2012. Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.19: Uncertainty Analysis. https://echa.europa.eu/documents/10162/13632/information_requirements_r19_en.pdf.
ECHA (European Chemicals Agency), 2014. Interface Between REACH and Cosmetics Regulations. https://echa.europa.eu/documents/10162/13628/reach_cosmetics_factsheet_en.pdf.
ECHA (European Chemicals Agency), 2017. Read-Across Assessment Framework (RAAF). https://echa.europa.eu/documents/10162/13628/raaf_en.pdf.
EFSA (European Food Safety Authority), 2017. Guidance on the use of the weight of evidence approach in scientific assessments. Efsa J. 15, e04971. https://doi.org/10.2903/j.efsa.2017.4971.
EFSA (European Food Safety Authority), 2018a. The principles and methods behind EFSA's guidance on uncertainty analysis in scientific assessment. Efsa J. 16 (1), e05122. https://doi.org/10.2903/j.efsa.2018.5122.
EFSA (European Food Safety Authority), 2018b. Guidance on uncertainty analysis in scientific assessments. Efsa J. 16 (1), e05123. https://doi.org/10.2903/j.efsa.2018.5123.
EFSA (European Food Safety Authority), 2019. Guidance on communication of uncertainty in scientific assessments. Efsa J. 17 (1), e05520. https://doi.org/10.2903/j.efsa.2019.5520.
EFSA (European Food Safety Authority) Scientific Committee, 2012. Scientific Opinion on exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC). Efsa J. 10 (7), 2750. https://doi.org/10.2903/j.efsa.2012.2750.
EFSA (European Food Safety Authority), World Health Organization (WHO), 2016. Review of the Threshold of Toxicological Concern (TTC) approach and development of new TTC decision tree. Efsa Support. Publ. 13 (3), 1006E. https://doi.org/10.2903/sp.efsa.2016.EN-1006.
Ellison, C.A., Blackburn, K.L., Carmichael, P.L., Clewell 3rd, H.J., Cronin, M.T.D., Desprez, B., Escher, S.E., Ferguson, S.S., Grégoire, S., Hewitt, N.J., Hollnagel, H.M., Klaric, M., Patel, A., Salhi, S., Schepky, A., Schmitt, B.G., Wambaugh, J.F., 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. https://doi.org/10.1016/j.yrtph.2019.01.016.
EMA (European Medicines Agency), 2018. Guideline on the Qualification and Reporting of Physiologically Based Pharmacokinetic (PBPK) Modelling and Simulation. . https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-physiologically-based-pharmacokinetic-pbpk-modelling-simulation_en.pdf.
Feigenbaum, A., Pinalli, R., Giannetto, M., Barlow, S., 2015. Reliability of the TTC approach: learning from inclusion of pesticide active substances in the supporting database. Food Chem. Toxicol. 75, 24–38. https://doi.org/10.1016/j.fct.2014.10.016.
Fisher, J.S., Turner, K.J., Brown, D., Sharpe, R.M., 1999. Effect of neonatal exposure to estrogenic compounds on development of excurrent ducts of the rat testis through puberty to adulthood. Environ. Health Perspect. 107 (5), 397–405. https://doi.org/10.1289/ehp.99107397.
Goebel, C., Troutman, J., Hennen, J., Rothe, H., Schlatter, H., Gerberick, G.F., Blömeke, B., 2014. Introduction of a methoxymethyl side chain into p-phenylenediamine attenuates its sensitizing potency and reduces the risk of allergy induction. Toxicol. Appl. Pharmacol. 274 (3), 480–487. https://doi.org/10.1016/j.taap.2013.11.016.
Hatherell, S., Baltazar, M.T., Reynolds, J., Carmichael, P.L., Dent, M., Li, H., White, A., Middleton, A.M., 2020. Identifying and characterizing stress pathways of concern for consumer safety in next generation risk assessment. Toxicol. Sci submitted.
Joint Research Centre, 2018. EURL ECVAM status report on the development. Validation & Regul. Acceptance Altern. Methods & Approaches 2018. https://doi.org/10.2760/818599.
Kalkhof, H., Herzler, M., Stahlmann, R., Gundert-Remy, U., 2012. Threshold of toxicological concern values for non-genotoxic effects in industrial chemicals: re-evaluation of the Cramer classification. Arch. Toxicol. 86 (1), 17–25. https://doi.org/10.1007/s00204-011-0732-z.
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. https://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. https://doi.org/10.1016/j.fct.2007.06.021.
Laufersweiler, M.C., 2012. Correlation of chemical structure with reproductive and developmental toxicity as it relates to the use of the threshold of toxicological concern. Regul. Toxicol. Pharmacol. 62 (1), 160–182. https://doi.org/10.1016/j.yrtph.2011.09.004.
Lobenhofer, E., Cui, X., Bennett, L., Cable, L., Merrick, A., Churchill, G., Afshari, C., 2004. Exploration of low-dose estrogen effects: identification of No observed Transcriptional Effect Level (NOTEL). Toxicol. Pathol. 32 (4), 482–492. https://doi.org/10.1080/01926230490483324.
Manwaring, J., Rothe, H., Obringer, C., Foltz, D.J., Baker, T.R., Troutman, J.A., Hewitt, N.J., Goebel, C., 2015. Extrapolation of systemic bioavailability assessing skin absorption and epidermal and hepatic metabolism of aromatic amine hair dyes *in vitro*. Toxicol. Appl. Pharmacol. 287 (2), 139–148. https://doi.org/10.1016/j.taap.2015.05.016.
Moxon, T.E., Li, H., Lee, M.Y., Piechota, P., Nicol, B., Pickles, J., Pendlington, R., Sorrell, I., Baltazar, M.T., 2020. Application of physiologically based kinetic (PBK) modelling in the next generation risk assessment of dermally applied consumer products. Toxicol. Vitr. 11 (December), 104746. https://doi.org/10.1016/j.tiv.2019.104746. [Epub ahead of print].
Munro, I.C., 1990. Safety assessment procedures for indirect food additives: an overview. Regul. Toxicol. Pharmacol. 12, 2–12. https://doi.org/10.1016/S0273-2300(05)80042-X.
Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G., 1996. Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. Food Chem. Toxicol. 34 (9), 829–867. https://doi.org/10.1016/S0278-6915(96)00049-X.
National Research Council, 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. The National Academies Press, Washington, DC. https://doi.org/10.17226/11970.
Obringer, C., Manwaring, J., Goebel, C., Hewitt, N.J., Rothe, H., 2016. Suitability of the *in vitro* Caco-2 assay to predict the oral absorption of aromatic amine hair dyes. Toxicol. Vitr. 32, 1–7. https://doi.org/10.1016/j.tiv.2015.11.007.
Organisation for Economic Cooperation and Development (OECD), 2005. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Series on Testing and Assessment. No. 34. ENV/JM/MONO(2005)14. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2005)14&doclanguage=en.
Organisation for Economic Cooperation and Development (OECD), 2007. Guidance Document on the Validation of (quantitative) Structure-activity Relationships [(Q)SAR] Models. Series on Testing and Assessment. No. 69. ENV/JM/MONO(2007)2. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)2&doclanguage=en.
Organisation for Economic Cooperation and Development (OECD), 2014. Guidance on grouping of chemicals. Series on Testing and Assessment, second edition. No. 194. ENV/JM/MONO(2014)4. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en.
Organisation for Economic Cooperation and Development (OECD), 2017. Chemical Safety Assessment Workflow Based on Exposure Considerations and Non-Animal Methods. Series on Testing and Assessment. No. 275. ENV/JM/MONO(2017)27. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2017)27&doclanguage=en.
Partosch, F., Mielke, H., Stahlmann, R., Kleuser, B., Barlow, S., Gundert-Remy, U., 2015. Internal threshold of toxicological concern values: enabling route-to-route extrapolation. Arch. Toxicol. 89 (6), 941–948. https://doi.org/10.1007/s00204-014-1287-6.
Pawar, G., Madden, J.C., Ebrell, D.J., Firman, J., Cronin, M.T.D., 2019. *In silico* toxicology data resources to support read-across and (Q)SAR. Front. Pharmacol. 10, 561. https://doi.org/10.3389/FPHAR.2019.00561.

Pinalli, R., Croera, C., Theobald, A., Feigenbaum, A., 2011. Thresholds of toxicological concern approach for the risk assessment of substance used for the manufacture of plastic food contact materials. Trends Food Sci. Technol. 22, 523–534. https://doi.org/10.1016/j.tifs.2011.07.001.

Przybylak, K.R., Schultz, T.W., Richarz, A.-N., Mellor, C.L., Escher, S.E., Cronin, M.T.D., 2017. Read-across of 90-day rat oral repeated-dose toxicity: a case study for selected b-olefinic alcohols. Comput. Toxicol. 1, 22–32. https://doi.org/10.1016/j.comtox.2017.02.002.

Raitano, G., Roncaglioni, A., Manganaro, A., Honma, M., Sousselier, L., Do, Q.T., Paya, E., Benfenati, E., 2019. Integrating *in silico* models for the prediction of mutagenicity (Ames test) of botanical ingredients of cosmetics. Comput. Toxicol. 12, 100108. https://doi.org/10.1016/j.comtox.2019.100108.

Rathman, J.F., Yang, C., Zhou, H., 2018. Dempster-Shafer theory for combining *in silico* evidence and estimating uncertainty in chemical risk assessment. Comput. Toxicol. 6, 16–31. https://doi.org/10.1016/j.comtox.2018.03.001.

Reilly, L., Serafimova, R., Partosch, F., Gundert-Remy, U., Cortiñas Abrahantes, J., Dorne, J.-L.M.C., Kass, G.E.N., 2019. Testing the thresholds of toxicological concern values using a new database for food-related substances. Toxicol. Lett. 314, 117–123. https://doi.org/10.1016/j.toxlet.2019.07.019.

EU-ANSA Research Cluster on 'Innovative 3Rs (Replacement, Reduction and Refinement of animal testing) approaches for the prediction of properties of chemicals, cosmetic ingredients, medicines, environmental contaminants and other regulated products'; ECHA, EFSA, EMA and the SCCS/SCHEER in consultation with the JRC-IHCP; April 2018. https://ec.europa.eu/health/sites/health/files/scientific_committees/docs/ansa_cluster_3r_en.pdf.

Russell, W.M.S., Burch, R.L., 1959. The Principles of Humane Experimental Technique. Methuen, London, pp. 1–238.

SCCS (Scientific Committee on Consumer Safety), 2010. Opinion on Hydroxyethyl-p-Phenylenediamine Sulfate. 23 March 2010. SCCS/1310/10, revision 12 July 2010. https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_019.pdf.

SCCS (Scientific Committee on Consumer Safety), 2012a. Opinion on p-Phenylenediamine. June 2012. SCCS/1443/11, revision 18 Sept 2012. pp. 26–27. https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_094.pdf.

SCCS (Scientific Committee on Consumer Safety), 2012b. Opinion on Toluene-2,5-Diamine and its Sulfate. June 2012. SCCS/1479/12, revision of 18 Sept 2012. pp. 26–27. https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_093.pdf.

SCCS (Scientific Committee on Consumer Safety), 2013. Opinion on 2-Methoxy-Methyl-p-Phenylenediamine and Its Sulfate Salt. 26 February 2013. SCCS/1491/12. https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_123.pdf.

SCCS (Scientific Committee on Consumer Safety), 2016. SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation 9th Revision. 29 September 2015. SCCS/1564/15, revision of 25 April 2016. https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_190.pdf.

SCCS (Scientific Committee on Consumer Safety), 2018. SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation 10th Revision. October 2018. SCCS/1602/18. pp. 24–25. https://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_224.pdf.

SCHEER (Scientific Committee on Health, Environmental and Emerging Risks), 2018. Memorandum on Weight of Evidence and Uncertainties. Revision 2018. 26 June 2018. https://ec.europa.eu/health/sites/health/files/scientific_committees/scheer/docs/scheer_o_014.pdf.

Schultz, T.W., Cronin, M.T.D., 2017. Lessons learned from read-across case studies for repeated-dose toxicity. Regul. Toxicol. Pharmacol. 88, 185–191. https://doi.org/10.1016/j.yrtph.2017.06.011.

Schultz, T.W., Rogers, K., Aptula, A.O., 2009. Read-across to rank skin sensitization potential: subcategories for the Michael acceptor domain. Contact Derm. 60 (1), 21–31. https://doi.org/10.1111/j.1600-0536.2008.01473.x.

Schultz, T.W., Przybylak, K.R., Richarz, A.-N., Mellor, C.L., Escher, E.S., Bradbury, S.P., Cronin, M.T.D., 2017. Read-across of 90-day rat oral repeated-dose toxicity: a case study for selected n-alkanols. Comput. Toxicol. 2, 12–19. https://doi.org/10.1016/j.comtox.2017.02.002.

Schultz, T.W., Richarz, A.-N., Cronin, M.T.D., 2019. Assessing uncertainty in read-across: questions to evaluate toxicity predictions based on knowledge gained from case studies. Comput. Toxicol. 9, 1–11. https://doi.org/10.1016/j.comtox.2018.10.003.

Thomas, R.S., Clewell 3rd, H.J., Allen, B.C., Wesselkamper, S.C., Wang, N.C.Y., Lambert, J.C., Hess-Wilson, J.C., Jay Zhao, Q., Andersen, M.E., 2011. Application of transcriptional benchmark dose values in quantitative Cancer and noncancer risk assessment. Toxicol. Sci. 120 (1), 194–205. https://doi.org/10.1093/toxsci/kfq355.

Thomas, R.S., Bahadori, T., Buckley, T.J., Cowden, J., Deisenroth, C., Dionisio, K.L., Frithsen, J.B., Grulke, C.M., Gwinn, M.R., Harrill, J.A., Higuchi, M., Houck, K.A., Hughes, M.F., Hunter 3rd, E.S., Isaacs, K.K., Judson, R.S., Knudsen, T.B., Lambert, J.C., Linnenbrink, M., Martin, T.M., Newton, S.R., Padilla, S., Patlewicz, G., Paul-Friedman, K., Phillips, K.A., Richard, A.M., Sams, R., Shafer, T.J., Setzer, R.W., Shah, I., Simmons, J.E., Simmons, S.O., Singh, A., Sobus, J.R., Strynar, M., Swank, A., Tornero-Valez, R., Ulrich, E.M., Villeneuve, D.L., Wambaugh, J.F., Wetmore, B.A., Williams, A.J., 2019. The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. Toxicol. Sci. 169 (2), 317–332. https://doi.org/10.1093/toxsci/kfz058.

Tluczkiewicz, I., Buist, H.E., Martin, M.T., Mangelsdorf, I., Escher, S.E., 2011. Improvement of the Cramer classification for oral exposure using the database TTC RepDose–a strategy description. Regul. Toxicol. Pharmacol. 61 (3), 340–350. https://doi.org/10.1016/j.yrtph.2011.09.005.

USEPA, 2014. Next Generation Risk Assessment: Incorporation Of Recent Advances In Molecular, Computational, And Systems Biology (Final Report). National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington D.C. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=286690.

van Ravenzwaay, B., 2017. The Threshold of Toxicological Concern for prenatal developmental toxicity in rats and rabbits. Regul. Toxicol. Pharmacol. 88, 157–172. https://doi.org/10.1016/j.yrtph.2017.06.008.

Wetmore, B.A., Wambaugh, J.F., Ferguson, S.S., Sochaski, M.A., Rotroff, D.M., Freeman, K., Clewell 3rd, H.J., Dix, D.J., Andersen, M.E., Houck, K.A., Allen, B., Judson, R.S., Singh, R., Kavlock, R.J., Richard, A.M., Thomas, R.S., 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol. Sci. 125 (1), 157–174. https://doi.org/10.1093/toxsci/kfr254.

Wetmore, B.A., Wambaugh, J.F., Allen, B., Ferguson, S.S., Sochaski, M.A., Setzer, R.W., Houck, K.A., Strope, C.L., Cantwell, K., Judson, R.S., LeCluyse, E., Clewell, H.J., Thomas, R.S., Andersen, M.E., 2015. Incorporating high-throughput exposure predictions with dosimetry-adjusted *in vitro* bioactivity to inform chemical toxicity testing. Toxicol. Sci. 148 (1), 121–136. https://doi.org/10.1093/toxsci/kfv171.

World Health Organization (WHO), 2017. Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterisation. second edition. Harmonization Project Document 11. http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf.

Yang, C., Barlow, S.M., Muldoon Jacobs, K.L., Vitcheva, V., Boobis, A.R., Felter, S.P., Arvidson, K.B., Keller, D., Cronin, M.T.D., Enoch, S., Worth, A., Hollnagel, H.M., 2017. Thresholds of Toxicological Concern for cosmetics-related substances: new database, thresholds, and enrichment of chemical space. Food Chem. Toxicol. 109 (1), 170–193. https://doi.org/10.1016/J.FCT.2017.08.043.

Zeller, A., Pfuhler, S., 2014. N-acetylation of three aromatic amine hair dye precursor molecules eliminates their genotoxic potential. Mutagenesis. 29 (1), 37–48. https://doi.org/10.1093/mutage/get053.

Regulation (EC) No. 1223/2009, 2009. of the European Parliament and the Council of 30 November 2009 on cosmetic products. Official Journal L 342, 0059–0209 22/12/.

Directive 2010/63/EC, 2010. of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Official Journal L 0033–0079. 276 of 20.10.. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:en:PDF.