A 10-step framework for use of read-across (RAX) in next generation risk assessment (NGRA) for cosmetics safety assessment

Camilla Alexander-White, Dagmar Bury, Mark Cronin, Matthew Dent, Eric Hack, Nicola J. Hewitt, Gerry Kenna, Jorge Naciff, Gladys Ouedraogo, Andreas Schepky, Catherine Mahony, Cosmetics Europe

PII: S0273-2300(21)00235-X

DOI: https://doi.org/10.1016/j.yrtph.2021.105094

Reference: YRTPH 105094

To appear in: *Regulatory Toxicology and Pharmacology*

Received Date: 25 November 2020

Revised Date: 12 July 2021

Accepted Date: 2 December 2021

Please cite this article as: Alexander-White, C., Bury, D., Cronin, M., Dent, M., Hack, E., Hewitt, N.J., Kenna, G., Naciff, J., Ouedraogo, G., Schepky, A., Mahony, C., Europe, C., A 10-step framework for use of read-across (RAX) in next generation risk assessment (NGRA) for cosmetics safety assessment, *Regulatory Toxicology and Pharmacology* (2022), doi: https://doi.org/10.1016/j.yrtph.2021.105094.

AlexWhite et al CRediT author statement

To be provided later

CRediT roles for EACH author. Please choose within this list: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing – original draft; Writing – review & editing.

A 10-Step Framework for Use of Read-Across (RAX) in Next Generation Risk Assessment (NGRA) for cosmetics safety assessment

Camilla Alexander-White[1], Dagmar Bury[2], Mark Cronin[3], Matthew Dent[4], Eric Hack[5], Nicola J Hewitt[6], Gerry Kenna[6], Jorge Naciff[7], Gladys Ouedraogo[8], Andreas Schepky[9], Catherine Mahony[10]

1. MKTox & Co Ltd, 36 Fairford Crescent, Downhead Park, Milton Keynes, Buckinghamshire. MK15 9AQ. UK. ckaw@mktoxconsult.com
2. L'Oreal Research & Innovation, 9 rue Pierre Dreyfus, 92110 Clichy, France
3. School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 AF, UK
4. Unilever, Safety & Environmental Assurance Centre, Colworth House, Sharnbrook, Bedfordshire. MK44 1ET, UK.
5. ScitoVation, Research Triangle Park, Durham, NC, USA
6. Cosmetics Europe, 40 Avenue Hermann-Debroux, 1160 Brussels, Belgium.
7. The Procter & Gamble Company, Cincinnati, OH, 45040, USA
8. L'Oreal Research & Innovation, 1 Avenue Eugène Schueller, Aulnay sous bois, France.
9. Beiersdorf, Unnastrasse 48, 20253 Hamburg, Germany.
10. Procter & Gamble Technical Centres Ltd, Reading, RG2 0RX, UK

Corresponding author: Cosmetics Europe, Avenue Hermann Debroux 40, 1160 Brussels, Belgium, cosmeticseurope@cosmeticseurope.eu

KEYWORDS: Next generation read-across (RAX), new approach methodology (NAM), next generation risk assessment (NGRA), cosmetics safety assessment, systemic toxicity, physiologically-based biokinetic modelling (PBK), caffeine, parabens.

ABBREVIATIONS: ADME, absorption, distribution, metabolism, excretion; CPR, Cosmetic Products Regulation; CSR, Cosmetic Safety Report; EU, European Union; IP, intraperitoneal; IV, intravenous; MACCS, Molecular ACCess System; MOA, mode of action; MOIE, Margin of Internal Exposure; NAM, New Approach Methodologies; NGRA, Next Generation Risk Assessment; PBK, physiologically-based kinetic; POD, point of departure; PBS, phosphate-buffered saline; RAX, Read-Across; RPF, relative potency factor; SCCS, Scientific Committee on Consumer Safety; SEURAT, Safety Evaluation Ultimately Replacing Animal Testing; SMILES, Simplified Molecular Input Line Entry Specification; TTC, Threshold of Toxicological Concern

Abstract

This paper presents a 10-step read-across (RAX) framework for use in cases where a threshold of toxicological concern (TTC) approach to cosmetics safety assessment is not possible. RAX builds on established approaches that have existed for more than two decades using chemical properties and in silico toxicology predictions, by further substantiating hypotheses on toxicological similarity of substances, and integrating new approach methodologies (NAM) in the biological and kinetic domains. NAM include new types of data on biological observations from, for example, *in vitro* assays, toxicogenomics, metabolomics, receptor binding screens and uses physiologically-based kinetic (PBK) modelling to inform about systemic exposure. NAM data can help to substantiate a mode/mechanism of action (MoA), and if similar chemicals can be shown to work by a similar MoA, a next generation risk assessment (NGRA) may be performed with acceptable confidence for a data-poor target substance with no or inadequate safety data, based on RAX approaches using data-rich analogue(s), and taking account of potency or kinetic/dynamic differences.

Introduction

Read-across (RAX) is defined as the use of relevant information from analogous substance(s) (the 'source' information) to predict properties for the 'target' substance(s) under consideration (ECHA, 2017). The concept of RAX in toxicological risk assessment has been around for at least two decades (Willett et al., 1998; Hanway & Evans, 2000; Kovarich et al., 2019; Krewski et al., 2020), as a way forward to contribute to a chemical safety assessment without the generation of new animal testing. Typically, RAX has looked at how chemistry can be used to predict toxicological properties (Cronin et al., 2017). The application of RAX is particularly pertinent where animal testing, for systemic toxicology and kinetics, is not legally possible as is the case for cosmetic ingredients in the European Union since 11 March 2013 (Laroche et al., 2018). In Europe, the Scientific Committee for Consumer Safety (SCCS) state in the 10th Notes of Guidance (SCCS, 2018) 'For the safety evaluation of cosmetic ingredients, all available scientific data are considered, taking into account the testing and marketing bans in force under Regulation (EC) No 1223/2009.' This includes all types of relevant scientific data, including the use of RAX, however in practice it has been challenging to position RAX into regulatory decision-making for human safety (Ball et al., 2016). As Patlewicz et al (2015) state 'Acceptance [*of read-across*] is undoubtedly thwarted partly by the lack of a systematic framework to characterise the read-across justification and identify the uncertainties particularly for complex regulatory endpoints such as repeated dose toxicity or prenatal developmental toxicity.' A framework for the positioning and application of RAX in cosmetics safety assessment is therefore much needed for assessing systemic toxicity without new animal data.

A number of developments now support the setup of such a framework: the EU-funded research programme, SEURAT-1, led to the creation of an exposure based mode of action (MoA) driven workflow for safety evaluation without generating animal data, including the option for RAX (Berggren et al., 2017); ECHA released a report in 2017 on their Read-

Across Assessment Framework (RAAF) (ECHA, 2017) for the purposes of EU REACH regulation; the OECD developed guidance on the grouping of chemicals (OECD 2014) and the US EPA have begun to incorporate general RAX (GenRA) approaches in the EPA CompTox Chemicals Dashboard (Shah et al., 2016; Helman et al., 2018; Helman et al., 2019; Thomas et al., 2019).

Since a workshop on NAM and RAX at the European Chemicals Agency in 2016 (ECHA, 2016), a number of RAX case studies have emerged in the literature (Mellor et al., 2016a, 2016b; Mellor et al., 2017; Przybylak et al., 2017; Schultz et al., 2017a, 2017b; Gelbke et al., 2018; Firman et al., 2018; Escher et al., 2019; Benfenati et al 2019; Luijten et al., 2020), describing tools, good practice and approaches to support human safety decision-making in a pragmatic and appropriately conservative way. These examples look at ways of demonstrating similarity of an analogue(s) to a target substance, using evidence of similarities with respect to chemical structure, physicochemical properties, metabolism and toxicokinetics, toxicodynamics, as well as similar structural alerts using predictive QSAR approaches for traditional toxicological endpoints. The ultimate intention is for these techniques to be useful in a regulatory context.

The National Academy of Sciences, Engineering and Medicine (NAS) 2017 also report on RAX 'Using 21st century science to improve risk-related evaluations' stating "*One approach for evaluating data-poor chemicals is to use toxicity data on well-tested chemicals (analogues) that are similar to the chemicals of interest in their structure, metabolism, or biological activity in a process known as read-across"* (see Figure 1 and figure legend for explanation).

[Insert Figure 1]

Our presentation of the concept of next-generation RAX in this paper, involves building on established approaches, with recognition of the specific challenges that face the cosmetic sector, in terms of ensuring that no evidence or information comes from new animal data performed for the purposes of the EU Cosmetics Regulation. The increasing availability of *in chemico* and biological *in vitro* NAMs in recent years now provides a significant improvement in the means to explore similarities and differences in metabolism, kinetics, toxicodynamics and biological activity between data-rich analogues and data-poor target chemicals (Escher et al., 2019; Krewski et al., 2020; Sauer et al., 2020).

The aim in NGRA, as with traditional risk assessment, is to derive a scientifically justifiable quantitative point of departure (POD) for a target substance and an endpoint that can inform either the derivation of a suitably protective Health Reference Value (HRV) and/or a quantitative risk assessment for a predicted or observed adverse health outcome. A POD for the target substance that has no or inadequate toxicology data is derived from a similar analogue substance by RAX. This POD is then used together with exposure data on the target substance to derive a margin of safety (MOS) and thus performing a quantitative NGRA, accounting for the level of confidence in the overall outcome.

With this in mind and the set of nine principles proposed by Dent et al (2018) to guide NGRA (Figure 2) it is apparent that a revised practical framework is needed for RAX, particularly to support transparent and structured risk assessment in a regulatory context, to enable the integration of both toxicokinetic and toxicodynamic based NAM data. Above all, we should remember Principle 1 from Dent et al (2018), that the overall goal is to assure human safety by performing an assessment that is relevant to humans.

[Insert Figure 2]

A Proposed 10-Step Read-Across Framework for Next Generation Risk Assessment

Figure 3 outlines a proposed 10-step framework structured in three tiers, showing the steps associated with a NGRA based on a RAX. RAX is a component of a NGRA extending the "traditional" read-across paradigm that has typically been applied, using chemical structures and properties, by the integration of further lines of evidence to generate and substantiate hypotheses relating to i) toxicodynamics (specifically the mode/mechanism of action (MoA)) and ii) toxicokinetics (relating to systemic bioavailability and metabolism). The approach benefits from being a flexible and iterative procedure, at times requiring reflection on the quality of read-across arguments in terms of associated levels of confidence, and how confidence can be increased by obtaining or generating data from NAMs. A variety of cheminformatics tools and *in vitro* assays can be used to inform on potential MoAs and kinetics in a tiered and iterative approach for both source (analogues) and the target chemical that is the subject of the RAX. This chemical and biological information is then used to support the overall weight of evidence RAX hypothesis that increases confidence by reducing uncertainty to an acceptable level, given a defined exposure scenario for the target chemical.

To facilitate the implementation of RAX in a regulatory context, a framework to organise and report the information is needed to enable transparent, reproducible and scientifically defensible decision-making. We detail a 10-step framework that, as can be seen from Figure 3, is an evolution of the basic ideas that emerged as an output from the EU-funded research project SEURAT-1 (Berggren et al., 2017)). The 10-step framework is a tiered approach to RAX that is exposure driven and MoA based. It is possible to exit the framework at the end of different tiers when confidence in the outcome is acceptable for a given exposure scenario. Two case studies (for propyl paraben and caffeine as target chemicals) applying

156 this 10-step framework accompany this manuscript (Ouedraogo et al., 2021; Bury et al.,
157 2021 respectively) and the reader is referred to these papers for learnings and practical
158 demonstration of the potential usefulness of the framework in supporting chemical safety
159 assessment. The next sections describe how to work through the framework step by step.

160 *[Insert Figure 3]*

Tier 0 - steps 1 to 4 of the 10-Step RAX framework

At the very beginning, one should consider the problem formulation and decision context, the degree of exposure and the information gaps that exist for the target chemical of concern. Clear and unambiguous problem formulation is required for NGRA (Embry et al., 2014; Cronin et al., 2019). This RAX framework is predominantly exposure driven and as Berggren et al (2017) describe, an early exit is possible during Tier 0 for chemicals where human exposure is very low and below a relevant threshold of toxicological concern (TTC) value, as defined on the basis of chemical structure and the known toxicity of chemicals sharing similar structural characteristics (EFSA, 2019; Yang et al., 2017).

This paper considers a RAX approach in cases where a TTC approach is not possible as exposure levels are higher than can be risk assessed using TTC, or the substance falls outside of the TTC application domain. The workflow of the Tier 0 process as shown in Figure 3, is discussed below.

Step 1: Identify exposure/use scenarios for target chemical

Assessment of the exposure to a cosmetic ingredient based on the product use scenario is a key part of cosmetics safety assessment (SCCS, 2018). In general, there is agreement that a tiered approach should be used for exposure estimates (Delmaar & van Engelen, 2006; Embry et al., 2014; Meek et al., 2011 – see Figure 4).

In Tier 0 of this RAX framework, the approach can range from conservative deterministic exposure estimates (as in the caffeine case study (Bury et al., 2021) derived using the maximum % of a chemical ingredient in product(s) together with information on maximum product usage, to a more sophisticated probabilistic modelling exposure estimate if necessary, according to the principles of the Scientific Committee on Consumer Safety 10th Notes of Guidance (SCCS, 2018), taking into account realistic habits and practices information of product use. Ingredient occurrence data using Industry Survey data and

consumer database information can also be used (as in the parabens case study (Ouedraogo et al., 2021)).

We are ultimately aiming to perform a risk assessment by calculating an acceptable margin of safety (MOS) where MOS = toxicological POD divided by an exposure estimate in the same units. Tiered risk assessment is an iterative process of refinement of both exposure estimates and toxicological hazard. If, at any tier, an acceptable margin of safety (MOS) cannot be demonstrated, the assessment moves to a higher tier where more data are generated to increase the level of confidence. However at lower tiers in the assessment, there is often more conservatism applied (Solomon et al., 2008). The safety assessment is finished if (at any tier of the approach) it has been demonstrated that the MOS is acceptable for the population under consideration, or if at the highest tier the risk is not acceptable and further refinements are not possible, then risk management measures such as restrictions for use must be put in place.

It is necessary to note that exposure in the present context mostly refers to an external exposure, i.e. external dose of the respective ingredient. So, in Tier 0 deterministic and probabilistic exposure evaluations are used to determine an external dose metric, usually for systemic toxicity endpoints in units of mg chemical/kg body weight/day.

However, only if a cosmetic ingredient enters the systemic circulation and reaches a target tissue or organ, can it be possible for a systemic adverse effect to occur. The internal exposure experienced by the body depends considerably on the route of exposure, including respective kinetic and metabolic differences. The majority of cosmetic products are applied via the dermal route and living skin is a barrier which can limit systemic exposure. Indeed, there are many factors that are important in the overall exposure assessment and their inclusion may be a consideration for further exposure refinement as needed as shown in Tier 1 of the framework. Knowledge about absorption, distribution, metabolism and excretion (ADME) (via skin, oral or inhalation routes) can provide an internal dose metric. Moreover, the application of physiologically-based biokinetic (PBK) modelling in Tier 2 may help to

target internal dose metrics (i.e. chemical in organs, blood, etc) in animals and humans. These kinetic data are considered in Tiers 1 and 2 of the proposed framework but for now, in Tier 0, we only consider external exposure.

Step 2: Identify molecular structure of target chemical

In the RAX approach presented here and in the accompanying case studies (Ouedraogo et al., 2021; Bury et al., 2021), there is a pre-requisite for the explicit definition of chemical structure of the target compound(s) (for which the RAX is to be applied) and that of the source compound(s) (analogues from which the data gap is filled via read across). The chemical structure should be defined explicitly using e.g. SMILES, INChi, IUPAC name and other relevant identifiers. Aspects such as stereochemistry, isomerisation and 2D structure should be clearly defined. Existing commercially available chemicals will have a Chemical Abstracts Service (CAS) number with the possibility of measured and predicted physicochemical properties. It is possible that traditional analytical chemistry approaches will enable the chemical structure of a truly novel chemical to be determined *de novo* as a starting point without it being a registered chemical. In certain cases, RAX may be possible for uncharacterised mixtures or UVCBs (Undefined Variable composition, Chemical/Biological) where only the major constituents are defined and there are gaps in the chemical similarity data, but where biological similarity data exists (Ryan et al., 2019; House et al., 2021). The application of RAX for mixtures such as botanical extracts is under development and promising progress is being made (Little et al., 2017; Vandermolen et al 2020).

In this step, structural features (such as molecular scaffolds or substructures, substituents, functional groups, isomers, tautomers, alkyl chain lengths) are identified for the target chemical that will enable the analogue search strategy in future steps. Any likely biotransformations and metabolites should be predicted and/or measured (in particular reactive metabolite formation) such that it can be hypothesised whether the parent chemical

(as in the case of the parabens (Ouedraogo et al., 2021)) or a metabolite (as in the caffeine case study (Bury et al., 2021)) is likely to act as the toxicant. The structures of major metabolites should also be known.

Step 3: Collate supporting data on target chemical and define data gap(s)

All attempts should be made to collate data for the target chemical (and major metabolites of interest) on physico-chemical properties; existing toxicology and NAM data; and absorption, distribution, metabolism and excretion (ADME) data. An appropriate literature and database search should be performed using the major authoritative sources of information and a search strategy adopted such as those described in IATA (Integrated Approaches to Testing and Assessment) case studies by the OECD (Van der Stel et al., 2021). A typical literature and data search might include sources such as those in Table 1. All evidence should be tabulated in a clear, systematic and logical form.

255 **Table 1** Useful sources of physico-chemical data, toxicological, ADME and NAM information.
256 Note this table is not exhaustive but includes major sources of information at the time of
257 publication.

Source	Weblink*
Physico-chemical Data	
US EPA CompTox Chemicals Dashboard	https://comptox.epa.gov/dashboard
ChemSpider	https://www.chemspider.com/
SciFinder	https://scifinder.cas.org/scifinder/view/scifinder/scifinderExplore.jsf
EpiSuite	https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface
Toxicological Data	
PubMed (National Center for Biotechnology Information)	https://pubmed.ncbi.nlm.nih.gov/
National Library of Medicine's Hazardous Substances Database Information	https://pubchem.ncbi.nlm.nih.gov/
European Chemicals Agency – REACH data	https://echa.europa.eu/
International Agency for Research on Cancer (IARC)	https://www.iarc.fr/
National Toxicology Programme (NTP) - USA	https://ntp.niehs.nih.gov/
OECD Screening Information DataSet (SIDS)	https://hpvchemicals.oecd.org/ui/Default.aspx
Japan – NITE-CHRIP Database	https://www.nite.go.jp/en/chem/chrip/chrip_search/systemTop
Japan Existing Chemicals Database (JECDB)	http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp
Cosmetic Ingredient Review (CIR)	https://www.cir-safety.org
EPA's Aggregated Computational Toxicology Online Resource (ACToR) – CompTox Dashboard	https://comptox.epa.gov/dashboard
QSAR – e.g. OECD Toolbox or the VEGA Hub	http://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm https://www.vegahub.eu/about-vegahub/
New Approach Methods	
PubChem and ChemIDPLus (National Center for Biotechnology Information)	https://pubchem.ncbi.nlm.nih.gov/ https://chem.nlm.nih.gov/chemidplus/
Multiple databases for computational toxicology, metabolism prediction, In silico profiling tools (eg within OECD Toolbox, ToxTree, DEREK etc)	Many databases listed and described in Pawar et al 2019; http://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm https://ec.europa.eu/jrc/en/scientific-tool/toxtree-tool
US EPA ToxCast Data	https://www.epa.gov/chemical-research/downloadable-computational-toxicology-data
Toxicogenomic tools: Comparative toxicogenomics database NTP DrugMatrix	http://ctdbase.org/ https://norecopa.no/3r-guide/drugmatrix

258 *correct at time of going to press

If, at the end of this process, the target chemical has relevant data gaps, so that no point of departure (POD) or health reference value for systemic toxicity can be derived, the data gap(s) and the problem formulation must be clearly defined. Potential analogues are then identified as in step 4, using a clearly described analogue search strategy.

Step 4: Analogue(s) a) Identify, b) collate existing data, c) determine similarity hypothesis (with or without MoA data)

Step 4 is a crucial step in RAX as it allows for the identification of suitable analogues (source chemicals) to be used as a surrogate(s) for the target chemicals, from which a toxicity benchmark can be read across for the target based on a justifiable hypothesis. The process of analogue identification is multifaceted and may require multiple iterations of refinement depending on the data found in database and literature searches. It is driven by aspects of chemistry and metabolism knowledge, chemical similarity, biological activity (mechanistic) concordance, toxicokinetic similarity, toxicodynamic similarity, availability of evidence to justify and support the similarity arguments and the availability and quality of data for the source compounds. Similarity hypotheses may vary in type, from simple structural analogues e.g. a common function(s) variation in carbon chain length (as in the RAX case study for parabens (Ouedraogo et al 2020)) to more complex arguments based around MoA or common metabolites (as in the RAX case study for caffeine (Bury et al., 2021)). It must be remembered that even the most elegant read-across hypothesis will be let down if there are insufficient high-quality data, so a pragmatic well reported evidence-based approach to analogue identification is critical.

a) Identify analogues

Potential analogues are currently identified using two-dimensional molecular similarity taking account of features defined from step 2 such as substructures and functional groups,

285 reactive chemistries as well metabolism. If a substance is part of a clear and obvious
286 chemical family, or homologous series (as with n-alkyl chain parabens for example
287 (Ouedraogo et al (2020)), then a structural analogue from the family may be found when
288 searching in a toxicological database or customised read-across tool such as the OECD
289 QSAR Toolbox, COSMOS DB or AMBIT. It is preferable, even when a homologous series is
290 known, to undertake a generic chemical similarity search using cheminformatics tools and
291 still proceed with care in inferring toxicity, as it is possible that unexpected toxicological
292 behaviour, possibly through biotransformaitions, can arise. Searches based on common
293 modes of action or metabolites would usually be undertaken when some prior knowledge
294 can be assumed or may be made available following *in silico* or even *in vitro* profiling. Such
295 information may also include structural alerts for chemical–biological interactions, as linked
296 to molecular initiating events (MIEs) in Adverse Outcome Pathways (AOP) (Cronin &
297 Richarz, 2017) as well as any NAM data that already exist.

298

299 *i)* *Similarity based on chemical class or homologous series based on common sub-*
300 *structures*: The simplest form of read-across is to utilise a direct structural
301 analogue e.g. a compound of the same chemical class or that shares a similar
302 molecular scaffold or predominant substructure with the same substituents or
303 functional groups (as is the case for parabens (Ouedraogo et al., 2021)).
304 Preferably there would be very limited structural differences between the target
305 and source compounds. The assumption here is that the common functional
306 groups will have the same mode of action and the parent substance drives
307 toxicity. Differences in potency within a homologous series are often as a result of
308 definable differences in chemical structure or ADME properties. This is a simple
309 technique founded in several decades of experience of considering classes of
310 High Production Volume (HPV) chemicals. It is powerful due to its transparency
311 in principle, but demonstration of the assumed mode of action driving toxicity may

not be trivial.

ii) *Chemical structure similarity*: Different cheminformatics tools may be used to search databases for appropriate source substances as analogues for the target. Computed similarity scores are popular because of their speed of use and ability to find closely related structural analogues which are not necessarily or obviously in the same chemical class or homologous series.

It is vital that the correct approach for assessing similarity is used. Generally these methods are provided to query available databases to find similar structures or may be accessed via websites such as ChemMine (https://chemminetools.ucr.edu/), however these can sometimes lack transparency. The similarity search method uses some means of reducing the chemical structure to a digital representation and then comparing these representations using an algorithm to compute a similarity score. The most frequent means of characterising a molecule is the creation of binary fingerprints, or bit strings, representing the presence or absence of individual sub-structural features e.g. a functional group. There are many methods of creating these fingerprints varying from general descriptions of organic chemistry, to functional groups relevant for toxicology (Cereto-Massagué et al., 2015). Mellor et al (2019) reviewed a number of the fingerprint methods as means to support read-across, with some showing better functionality. Also, fingerprints based on toxicologically relevant functional groups (e.g. ToxPrint chemotypes, https://toxprint.org/) often perform well. A method is also required to determine 'relative similarity' between two molecules. There are many methods to do this, a simple approach that is widely used is the application of the Tanimoto Index which assesses the proportion of the overlap in fingerprints between two molecules (Bajusz et al., 2015). It is essential to note that the similarity between two molecules is a function of the fingerprint that is used. Different fingerprints will find different

analogues, and the Tanimoto coefficients will vary according to the fingerprint and the information it contains. Often the use of chemical structure-based similarity can yield quite a number of potential 'similar' substances or substances with different toxicological profiles from the target compound. The search can be refined by identifying the molecular scaffold or predominant structural features with required functional groups and similar physico-chemical properties and searching the relevant database by substructure (Wu et al., 2010, Lester et al., 2018).

It should also be remembered that source analogues with limited toxicology data, and particularly for the endpoint(s) of interest required for the target chemical, are not viable candidates for further RAX consideration, so informatics protocols for searching for this dependency early on are useful. Once this process is complete, considerations around the nature of the analogues returned should be made e.g. regarding core structural components, reactive groups, and structural variations that are or are not hypothesised to impact the toxicological outcome.

iii) *Physico-chemical property similarity*: Similarity can also be assessed in terms of physico-chemical properties and the associated data. Except in specific circumstances (e.g. pKa relating to skin corrosion), which are unlikely to be relevant for chronic human health effects, physico-chemical properties in themselves are not the basis of a read across argument. However, they will provide strong supporting evidence for toxicological similarity for similar chemical structures and are essential to assist in the determination of differences in toxicokinetics and potency since physico-chemical properties may affect bioavailability and consequently biological responses observed *in vitro* or *in vivo*.

A number of types of physico-chemical data could be included to assess similarity, for instance those that are requested in the SCCS 10th Notes of Guidance (2018): appearance/physical state, molecular weight, purity/impurity profile, presence of isomers, solubility in water, oil-water partition coefficient (log Pow), homogeneity, stability and identity of degradation products. Other basic properties would include: melting point, boiling point, density, vapour pressure, flash point, oxidising properties, pKa dissociation constant, pH in water, viscosity, polar surface area and number of rotatable bonds. Any special considerations can be added if necessary, e.g. for polymers or nanoforms of particulate materials. Other parameters such as hydrogen bond donors and hydrogen bond acceptors, which are part of the Lipinski Rule of Five (Lipinski et al., 2001) for predicting bioavailability, are useful to compare.

iv) *Similarity of common metabolite or degradant*. Similarity between molecules can be assumed if they elicit the same metabolite or a common degradant (as in the caffeine case study (Bury et al., 2021)). In this case the read-across is usually performed considering a principal metabolite, assuming toxicological data exist for it (Ball et al., 2014). Often studies using this approach will utilise relatively similar starting molecules, but it does also allow for compounds that are not close structural parent analogues to be considered, when they yield the same metabolite. This approach forms one of the cornerstones of the read-across scenarios in ECHA's Read-Across Assessment Framework (RAAF) (ECHA, 2017). Similarity of this type usually requires prior knowledge and detailed technical considerations, in addition, experimental evidence of extent and rates of metabolism (from *in vitro* studies, PBK or human evidence) may be required to strengthen the RAX argument.

v) *Biological similarity*: Biological, or toxicological, similarity can be used either to identify analogues or support similarity hypotheses. Similarity, in terms of a common toxicological mode/mechanism of action (MoA) for target and analogue(s) may be assumed from commonality of chemical structure. Chemical-specific MoAs can be considered together with a chemical-agnostic biological adverse outcome pathway (AOP)) (Ankley & Edwards, 2018). In an AOP there may be known key mechanism(s) associated with an identifiable molecular initiating event (MIE) (Cronin and Richarz, 2017). This could include gene expression markers or receptor based mechanisms, for example. Generating a hypothesis on a common MoA and also demonstrating similar patterns for similar chemicals may be achieved through *in silico* profilers. For instance Mellor et al (2016a) developed profilers for nuclear receptor ligands as an upstream mechanism of action associated with hepatic steatosis (fatty liver disease). Another example of a common MoA is the blocking of A1-adenosine receptors by methylxanthines, as shown in the caffeine case study (Bury et al., 2021).

Biological similarity can also be supported directly from experimental data where it exists and this may be combined with physico-chemical properties to increase confidence e.g. the Chemical–Biological Read-Across (CBRA) approach (Low et al., 2013). The 'Generalized Read-Across (GenRA)' approach developed by the US EPA's National Center for Computational Toxicology (NCCT) is an approach where in addition to chemical similarity, NAM bioassay data from the ToxCast programme have been used to cluster substances based on biological similarity (Shah et al., 2016; Patlewicz et al., 2017; Helman et al., 2018; Helman et al., 2019). It should however be noted that special consideration must be given to the potential issues with 'hit calls' from flawed dose response curves. Blackburn et al (2020) have proposed an initial framework for use of ToxCast data into a RAX safety assessment that takes account of this.

Following the identification of suitable analogues for the target substance, step 4b will be undertaken to retrieve appropriate toxicological data. These data should also be tabulated in a clear, systematic and logical form and analysed for concordance with the target and other source substances. If data are lacking or available data are insufficient in terms of study quality, other analogues may have to be sought as described above which may require attempting other means of searching for and selecting appropriate analogues. This is therefore an iterative process based on the analysis of the analogue data set.

b) Analogue – collate existing data

Iteratively, as analogues are being identified and narrowed down for comparison with the target compound using the chemical/biological similarity approaches as described above, searches for toxicological data (for the closest analogues) should be performed from all available sources (e.g. in Table 1). Data on the selected analogues may be available in global public data sources, such as from PubMed or authoritative reviews, such as the Cosmetics Ingredients Review (CIR) website, the European Chemicals Agency (ECHA) REACH database (as publication permits), or data may be available in company archives. It is relevant to collect all toxicological endpoint data for target and analogues, e.g. on the range of endpoints as would be included in an evaluation according to the SCCS 10th Notes of Guidance (2018), not just the study data type that is relevant to the target chemical data gap, as other data can also be used to substantiate biological and toxicological similarity. Ideally, the study quality should be reviewed and reported for all source analogues, most commonly undertaken as per the methods of Klimisch et al (1997) or possibly using the approach developed as a webtool for both legacy *in vivo* data and *in vitro* data in the SciRAP project (http://www.scirap.org/).

In order to address the toxicological data gap defined in Step 3, good quality systemic toxicology studies (e.g. 28-day, 90-day, 2-year bioassay, reproductive, developmental study data etc) are needed (ideally performed to OECD guidelines and to GLP, but these conditions may not always be available) to determine a toxicological POD for at least one close analogue. If there is a homologous series of analogues, it is necessary to collect all toxicology data for all analogues. There may be a good quantitative dose response, from which a no (or low) observed (adverse) effect level (NOAEL/LOAEL) or a benchmark dose (BMD) can be derived; or no effects may be seen in the study for the closest analogue and the top dose in the study acts as the NOAEL. The analogue selection and choice of data must be justified in a structured and rigorous report (similar to the approach described in Schulz et al 2015), as being sufficiently similar to the target, such that a similar MoA leading to the same degree and type of toxicity can be assumed for both. Given, the availability of good data, the POD for the analogue can then be read across for the target, with a certain level of confidence. However, the possibility of differential potencies must be considered and will be discussed below in terms of substantiating the RAX and assessing confidence in an outcome.

c) Determination of similarity hypothesis

It is important early on in a RAX to give some thought to the overall hypothesis for the similarity between the target and source chemicals. It is also important to consider any impacts on the selection of analogues and accompanying data from any differences that may arise from the route and duration of exposure for target and analogue substances. In this regard it may be helpful to consider the scientific basis and definition of the RAX scenario on the basis of chemical similarity, metabolism and mode of action as follows (Berggren et al., 2015; ECHA, 2017):

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

In one of our accompanying case studies, a search for chemically similar analogues and existing knowledge on caffeine metabolism led to the decision to use the major caffeine metabolites as source chemicals for the target caffeine (Bury et al., 2021). In the other case of parabens, the search for chemically similar analogues led to a focus on the homologous series of short-chain n-alkyl parabens, with the use of physico-chemical properties to help define the boundaries of similarity (Ouedraogo et al., 2021). It is essential that the similarity hypothesis is pragmatic, can be supported by sufficient evidence to make it acceptable and fit for purpose, and crucially there are sufficient high-quality toxicity data for the source molecules.

To summarise, when selecting read-across source (analogue(s)) for the target (chemical of interest), it is necessary to address the following considerations:

- Chemical structure, physicochemical, reactivity and metabolism properties
- Whether *traditional* good quality, quantitative toxicology data are available
- Whether *in vitro* data and/or NAM data are available
- Hypothesis of potential mode/mechanism of action common to the source and target chemicals in terms of the scenario to be assessed (as per categories I to IV above)

As one works through the Tier 0 process, it is likely that in the first instance a ‘category’ of similar analogues will be formed and then as more information on toxicological data and

biological similarity becomes available this iteratively focuses the analogue selection to either one analogue or a small category of analogues with suitable good quality quantitative data to derive a POD for the data gap defined in Step 3.

At this point in the process it should be considered whether there are sufficiently similar analogue(s) available with good quality toxicological data from which to move to Step 8 and perform a RAX to using a POD for the target based on an analogue. If exposure is reasonably low and the POD is predicted to be high from the RAX, it may be possible to derive an acceptable MOS in Step 9, with the level of confidence in the POD prediction (low, moderate or high) accounted for in the final assessment in Step 10. The derivation of a POD for the risk assessment using RAX is explained below in Step 8. If the level of confidence is not acceptable at this point in the framework, further data can be generated in Tier 1 related to bioavailability and to substantiate biological similarity, thus increasing confidence in the RAX hypothesis.

Tier 1 – Bioavailability/kinetics and MoA as relevant to the RAX

At the end of Tier 0, it may be the case that the level of confidence in the RAX derived POD for the target is not sufficiently high and additional safety factors need to be used in the risk assessment to account for uncertainties, either in potential differences between target and source kinetics or mode of action. In some instances, exposure may be so low and the MOS sufficiently high that the risk assessment is acceptable but given a higher exposure estimate, the resulting MOS may not be considered acceptable and further refinement may be needed. In Tier 1 such additional data can be generated using human relevant NAM approaches.

Step 5: Systemic bioavailability/ADME of target chemical and analogues

Legacy toxicity data most often come from orally dosed studies in animals. The exposure scenarios for a cosmetic are most often via dermal application. A default assumption of

similar absorption both orally and dermally is the starting point for the safety assessment but it is more realistic to determine or predict both the oral and the dermal absorption between target and analogue(s), so that a refinement of the POD can be made or uncertainty factored into the risk assessment to account for any relevant toxicokinetic differences in a proper manner.

Dermal penetration data can be generated for cosmetic ingredients by *in vitro* techniques (according to OECD guideline method 428 and criteria according to the SCCS Notes of Guidance), using excised human skin obtained ethically from cosmetic surgery operations, to estimate the systemic bioavailability i.e. a surrogate for an internal dose. Skin penetration could in theory be estimated *in silico* and a model has been developed for fragrance materials as described in Shen et al (2014) to predict flux (J_{max}). However, *in silico* predictions require further development for a broader range of chemicals. Recently, Hewitt et al (2020) described dermal absorption datasets for 56 chemicals, and it is hoped that these data can improve the predictivity of *in silico* modelling in the future.

Gathering as much evidence as possible on the similar and differential ADME properties for both target and analogues will help to inform how confident the POD prediction is for the target, whether adjustment is required and to ensure an appropriate level of conservatism is factored into the risk assessment for differential toxicokinetics between target and analogue and between routes of exposure. *In vitro* ADME parameters could be available in the literature (e.g. as per sources in Table 1), or from using *in silico* tools (a range of ADME data sources is listed in Pawar et al. 2019). Data may need to be generated to further justify the analogue selection, for example, on extent of protein binding, comparative enzyme kinetics (e.g. esterase activity for parabens), rate and extent of clearance (e.g. metabolism in liver), comparative gut permeability, transporter effects. These are just some of the possible bespoke ADME studies that could be performed and the data should be described in accordance with an OECD guideline or other appropriate guidance, include an analysis of

554 data quality and be expressed in clear and readable tabular form to enable an assessment
555 of similarity in bioavailability between target and analogue(s).

556 It could also be useful to use PBK modelling to estimate the internal dose metrics for route to
557 route comparisons, although many of the aforementioned ADME parameters will be needed
558 to do this. It may also be possible to derive PBK parameters using RAX approaches where
559 there are data gaps (Ellison & Wu, 2020), as well as POD endpoints. When kinetics are fully
560 accounted for, it allows for reduction of the uncertainty factors in relation to inter-species
561 toxicokinetics and PBK modelling has been used effectively in this regard in the case study
562 on parabens (Ouedraogo et al., 2021). More is described on this in Tier 2, Step 7b below.

563

564 **Step 6: Supporting a Similar Mode/Mechanism of Action hypothesis**

565 A mode or mechanism of action (MoA) hypothesis is the ideal starting point in deriving the
566 initial similarity hypothesis for the analogue selection of a target. This can be a difficult step
567 for cosmetic ingredients that are typically of no or low toxicity, for example there may be
568 observations such as body weight loss or gains, where a MoA is difficult to establish. In
569 situations where a MoA is not able to be defined well, it may only be possible to consider the
570 concordance between in vivo and in vitro data.

571 Based upon the findings of the chemical and biological activity similarity searches in Tier 0,
572 one can begin to form hypotheses for how the target and analogue(s) could act via similar
573 mechanisms and MoA in the body upstream of a serious adverse outcome, such as
574 reproductive or developmental toxicity, liver damage, cancer etc. Evidence on a MoA could
575 help to provide further justification for the RAX. For example, the adenosine receptor driven
576 MoA for a common metabolite in the accompanying caffeine case study (Bury et al., 2021).

577 As well as using evidence from existing *in vivo* studies that a similar MoA is at play for the
578 target and analogue(s), one could look for further evidence from NAMs to strengthen and
579 support the hypothesis. Toxicogenomics data are likely to be a key resource here and a

good example of its application to confirm similarity in MoA is provided by Chen et al., (2020). Liu et al (2019) describe a vision beyond 2020 for how toxicogenomics and big data approaches could be applied in RAX to underpin similarity hypotheses and toxicogenomics data are also used in the case study on parabens (Ouedraogo et al., 2021), providing an untargeted approach to inform on MoA and similarity. Toxicogenomics analyses can be performed using *in vitro* models such as tissue slices, primary hepatocytes, cell lines, 3D tissue models, stem cells, organoids or organ-on-a-chip models (De Abrew et al., 2015, 2019; Jiang et al., 2019; Liu et al 2019; Moroni et al., 2020; Punt et al., 2020) as well may be existing from *in vivo* studies. Van Ravenzwaay et al (2016) demonstrated when forming a category of analogues using physico-chemical properties for the target substance MCCP, a herbicide, how metabolomics data from the two closest analogues may also be used to substantiate the RAX and avoid the need to perform a 90-day study. In this case, metabolomics data were available for all three substances from 28-day toxicity studies, and valid 90-day toxicity studies for the two source substances, so legacy *in vivo* data were used to prevent the need for a new 90-day animal study for the target substance. Similarly, Sperber et al (2019) have actively demonstrated the principles of using metabolomics data to support RAX for a REACH submission for 3-aminopropanol (3AP), based on read-across from 2-aminoethanol (MEA).

A connectivity map (CMap) approach has been utilised by De Abrew et al (2019) where transcriptional profiling data, obtained *in vitro* from a number of different cell lines, is used to augment a RAX initially based on chemical properties and help assure that small differences in chemical structure between target and analogues do not have significant biological consequences. De Abrew et al describe two case studies; one for alkylphenols and another for diaminobenzenes. In each case, they used the data to support why some analogues were more suitable for RAX than others, due to biological activity similarity profiles. We have also used the CMap approach in our accompanying parabens case studies to affirm analogue identification and underpin the choice of analogue POD (Ouedraogo et al., 2021).

All be it complex in its analysis, such an approach provides a broad biological coverage although a targeted panel of ligands assays also proved useful in the parabens case studies for a deeper understanding of effect at the molecular level. It can be expected that as high content data streams continue to grow, spanning a larger chemical space, it may be increasingly possible to derive analogues based on functional similarity rather than strict analogy of chemical properties.

At this point at the end of Tier 1, as above, it could be possible to exit the framework and move to Steps 8-10. However, if available data qualitatively support similarity but provide insights into quantitative differences in the kinetic or biological activity profiles it may be necessary to perform targeted testing and biokinetic refinements in Step 7a.

Tier 2 Targeted MoA testing and biokinetic refinement to support RAX

Where increased confidence is needed in the kinetics or MoA hypothesis and potency derivation it should be investigated as to whether NAM data could be generated at Tier 2. The selection of *in vitro* assays that are relevant to run in targeted Tier 2 testing would be informed by the analysis performed in Tier 1. Also, refined estimates of internal dose using PBK modelling may help to reduce uncertainty and achieve a more realistic MOS.

Step 7a: Targeted testing using NAM biological assays to strengthen hypotheses

Based upon the information collated and generated in Tier 0 and Tier 1, customised and targeted testing may be needed to provide further evidence that supports a common MoA hypothesis for the target substance and the source analogue(s).

For example, it may be hypothesised that the target substance and analogue(s) all exert common toxicity through a receptor-driven mechanism of action associated with a specific adverse outcome. This is the case in the caffeine case study based on adenosine receptor

antagonism and nervous system/cardiovascular effects (Bury et al., 2021). There have been other promising uses of biological assays for support of a RAX paradigm. Coulet et al (2019) used a battery of *in vitro* bioassay data for nitrogen-containing polycyclic aromatic hydrocarbons (PANHs or aza-arenes), which are toxicologically data poor chemicals, to perform RAX from more data-rich polycyclic aromatic hydrocarbons (PAHs). The hypothesis tested was that PANHs are more potent inhibitors of aryl hydrocarbon receptors (AhR) than PAHs, and hence potentially more toxic. So as to avoid a significant amount of new animal testing, the concept of MIEs was used to explore the mechanisms responsible for carcinogenicity with PAHs. The problem formulation was to address whether PANHs are more potent carcinogens than PAHs. The MIE was described as the binding of the PAH benzo[a]pyrene (B(a)P) to the transcription factor AhR followed by induction of cytochrome P450 (enzyme) genes and subsequent B(a)P biotransformation into DNA reactive metabolites, DNA-adduct formation, mutations, and ultimately cancer pathologies. Assays included Chemical Activated Luciferase gene expression (CALUX) gene assays for reporter cell lines were used: the Estrogen Receptor alpha and beta (ERα, ERβ), the Androgen Receptor (AR) and the Aryl hydrocarbon Receptor (AhR); standard Ames test and flow-cytometric micronucleus tests for genotoxicity and the Phospho-ɣ H2AX activation test (Cellomics). These data showed that PAHs and PANHs could be best grouped in terms of the number of rings in their structure and molecular size; substances with 5 aromatic rings had similar biological properties to each other, different from 4 ringed and 3 ringed structures.

Targeted testing was also useful in the two case studies that accompany this paper. Paraben gene expression data pointed to *in vitro* estrogen receptor assays as being potentially useful to assess similarity of short-chain parabens, so Toxcast ER data across multiple assays was used to inform potency by virtue of the reported AC10 values across this homologous series (Ouedraogo et al., 2021). For the caffeine case study, CMap hits showed CNS and CVS activity, and published ligand affinity Ki values for relevant targets

were used to inform potency of methylxanthines (Bury et al., 2021). In both instances the mechanistic index, i.e. AC10, Ki values were justified and considered in relation to their proximity to the internal exposure estimate informed in Step 7b. As at all Tiers of the framework, the reporting and presentation of the data such that it is transparent and understandable is of critical importance. In both case studies the targeted testing revealed differences in biological potency in the target versus the analogue(s) from which it was possible to derive relative potency factors (RPFs) for application to Step 9.

Often, there may be multiple biochemical and biological mechanisms at play in the generation of an adverse health effect, so comparing the action of substances at the receptor binding level is just one piece of molecular evidence to substantiate that the substances act similarly through a molecular initiating event (MIE) but is not the total picture of biological adversity in an intact organism. This is especially true for substances that modulate endocrine systems, and where modulation may or may not lead to adversity. In the parabens case it was conservatively assumed that the parent substance is the driver of toxicity but targeted testing in an Estrogen, Androgen, Thyroidogenic, Steroidogenic (EATS) panel, in the presence of S9 incubations, showed a decrease in bioactivity suggesting, in fact ready metabolism to yield inactive metabolites (Ouedraogo et al., 2021).

Step 7b: Biokinetic refinements of target chemical and analogues

PBK modelling contributes to various aspects of NGRA. For instance, it can provide an estimate of the internal concentration in humans of the target substance at the plasma/organ level. Similarly, it is also possible to use PBK modelling to determine relevant internal tissue doses *in vivo* and then use this information to set doses for *in vitro* testing that would compare well with what the target cells are exposed to *in vivo*. Campbell et al (2015) used an approach such as this for the parabens. There are various exposure metrics that can be considered e.g. Cmax, AUC which are best informed by mechanistic considerations. In both

685 the parabens and caffeine case study the MoA was a direct receptor mediated effect and so
686 maximum free concentration in blood (Cmax) was relied upon (Ouedraogo et al 2020, Bury
687 et al., 2021). Where total dose over time is a consideration, the Area Under the Curve (AUC)
688 may be a better choice. Considerations over the dose metric that is most appropriate to use
689 can be found in Groothuis et al., (2015).

690 In the reverse manner it is also possible to use PBK models to perform quantitative *in vitro* to
691 *in vivo* extrapolation (QIVIVE) (Punt et al., 2020). For example, intracellular nominally
692 effective concentrations (e.g. an EC10, a concentration that yields a 10% increase in effect)
693 of a chemical in *in vitro* experiments can be derived. This effective dose in cells can then be
694 used to compare and relate to an *in vivo* dose at a target organ or molecular target *in vivo*.
695 Escher et al (2019) and Punt et al (2020) explain further the details of PBK approaches and
696 how they can help in grouping chemicals and NAM-based NGRA.

697 In the parabens and caffeine case studies that accompany this paper, PBK modelling is
698 used to determine the estimated in human blood/plasma (µg/L) such that this could be
699 related to an internal blood/plasma POD concentration from a toxicity study in animals for
700 the analogue. In the parabens case it was also applied to determine the plasma
701 concentration associated with the analogue NOEL (Ouedraogo et al., 2021). PBK modelling
702 can be used as a means to compare kinetics of target and analogue substances in a
703 category (as per the parabens case study (Ouedraogo et al., 2021)), and to compare kinetics
704 between routes of exposure and between species (as per the caffeine case study (Bury et
705 al., 2021)).

706 A range of software tools are available and input parameters are needed to build, validate
707 and use substance-specific PBK models. Madden et al (2019) reviewed an extensive list of
708 different sources of software and data for PBK modelling and highlighted their increasing use
709 in the pharmaceutical sector over the past 30 years. It is important that PBK models are
710 robust, scientifically credible and reproducible. In 2010, the World Health Organisation stated
711 that a PBK model is 'a model that estimates the dose to target tissue by taking into account

712 the rate of absorption into the body, distribution and storage in tissues, metabolism and
713 excretion on the basis of interplay among critical physiological, physico-chemical and
714 biochemical determinants' (WHO, 2010). WHO in 2010 developed guidelines for reporting
715 PBK models and the European Medicines Agency (EMA, 2016) have also published
716 guidelines on how a PBK model and how it has been parameterised should be reported.
717 Tan et al (2020) have proposed a structured reporting template to support the
718 communication and regulatory acceptance of PBK models, as they are complex
719 multiparametric models, where parameter selection can have a significant impact on
720 outcome. Most recently, the OECD have published a new guidance document for the
721 characterisation, validation and reporting of PBK models for regulatory purposes (OECD
722 2021). Therefore, such guidelines are relatively detailed and PBK modelling is a specialist
723 endeavour. Verifying model predictions (e.g. using existing in vivo data) and understanding
724 sensitivity towards different parameters is a critical component of the endeavour. A degree of
725 error in the PBK simulations may be acceptable if the uncertainty and direction of the
726 inaccuracy is described and its impact is considered relative to the protection goal. Some
727 examples of useful models available in the literature are for bisphenol A, 2-butoxyethanol,
728 methylene chloride, perchlorate, D5 and phenoxyethanol (Clewell et al., 2008; Fisher et al.,
729 2011; Troutman et al., 2015; McMullin et al 2015).

730 PBK models can also be built/verified using *in vivo* data and parameters on analogues to
731 inform the kinetics of a target substance that does not have *in vivo* kinetic study data and in
732 cases where no new kinetic data can be generated in animal models, such as for a cosmetic
733 ingredient in the EU. Building models using *in vivo* data on 'PK analogues' and then using
734 chemical-specific parameters for the target chemical allows for description of kinetics for the
735 target chemical. Ellison & Wu (2020) tested out this approach for caffeine (see more details
736 in Bury et al, 2020).

737 Increasingly human biomonitoring data are being generated in the general population to
738 demonstrate internal exposure to consumer product substances (e.g. the Human

739 Biomonitoring Project for the EU (HBM4EU) project). However, in general the exact external
740 exposures and sources are not known that lead to the observed biomonitoring measures
741 and thus it is difficult to derive meaningful conclusions. Nonetheless, such data can help
742 verify PBK estimates and determine that internal exposures are generally low and can
743 support a risk assessment conclusion.

744

745 **Step 8: Performing a RAX to derive a POD**

746 A RAX can be iteratively performed to yield a POD at the end of either Tier 0, Tier 1 or Tier
747 2. Most often toxicology studies are performed by the oral route in animals. For cosmetics
748 safety assessments, the dermal route is usually the most important, since most cosmetic
749 products are applied on the skin. However, the oral route is important for oral care products
750 such as toothpaste and mouthwash (which fall under cosmetics products in the EU).
751 Inhalation is a route of exposure to consider for ingredients which are in spray and aerosol
752 products, and sometimes there is a specific inhalation toxicology POD available.

753 At Tier 0, in using a RAX approach, the POD for the effect of concern from data on the most
754 chemically and biologically similar analogue is used directly as the same POD for the target
755 substance. This is the simplest conclusion to draw i.e. that similar chemical properties lead
756 to similar toxicity. The level of confidence as to whether the POD is likely to be the same, or
757 at least a conservative estimate for the target chemical needs to be considered. What is the
758 chance the POD could be substantively lower for the target than the POD of the analogue? It
759 is useful to consider not just toxicological No Observed Effect Levels (NOAELs) but also
760 Lowest Observed Effect Levels (LOAELs) and also dose response curves (if available) to
761 assess the potency of the analogues. An additional uncertainty factor may be required if
762 there is low confidence in the POD being relevant for the target and it may cause one to
763 reconsider analogue selection. In Tier 1, further evidence from ADME experiments or MoA
764 NAM data may provide a higher degree of confidence that the target chemical is similar in
765 effects and potency to the analogue. The target could be less or more systemically

766 bioavailable or more or less potent than the analogue in some assays etc. The POD may
767 well be taken as the same value but with higher confidence that the target will not be more
768 toxic than the analogue.

769 In Tier 2, if an acceptable MOS is not achieved with confidence after Tier 1, it may be
770 possible in Tier 2 to refine the risk assessment even further by using an internal
771 blood/plasma concentration metric for both the POD metric and the systemic exposure dose,
772 as relevant to humans using PBK modelling. Also, with targeted testing, more data from *in*
773 *vitro* assays could yield relative potency information, and a relative potency factor (RPF)
774 could be used to adjust the POD up or down, depending on how the target behaves relative
775 to the analogue. A RPF can be used in the final risk assessment, as is the case in the
776 caffeine case study (Bury et al., 2021). See also the parabens case study by Ouedraogo et
777 al (2020) as to how the POD was derived in this case.

778 **Step 9: Performing an MOS evaluation**

779 Once an estimate of exposure (as either an external dose (mg/kg/day) or internal dose
780 metric (e.g. µg/L blood)) and a POD for the target chemical have been derived from RAX
781 one can calculate a margin of safety (MOS) by dividing the POD by the corresponding
782 exposure metric (i.e. in the same units) for the ingredient in a product use scenario.

783 MOS = POD (read across from the most suitable analogue)/exposure estimate for the target

784 Whether the MOS is acceptable depends on the scale of uncertainty or level of confidence
785 as to how realistic the exposure estimate is, the level of confidence in the POD using RAX
786 and whether there are differences expected in the toxicokinetics and toxicodynamics
787 (TK/TD) between species, between human individuals or between the target and analogue.

788 At the end of Tier 0, comparison of an external dermal applied dose (mg/kg/day) with an
789 external intake oral dose POD (in mg/kg/day) from a toxicology study is a conservative
790 approach, as dermal absorption of a cosmetic ingredient is in reality lower than oral
791 absorption, due to the skin being an excellent protective barrier. In Tier 0, measured data on

792 dermal absorption of the chemical into the body does not exist. At the end of Step 4 in Tier 0
793 it could be possible to be confident in a POD that is based on good data from an analogue
794 and where a similarity analysis indicates that the predicted POD is conservative for the
795 target. It is possible that an acceptable MOS can be achieved after Tier 0 using a simple
796 worst-case assessment. In this case, where confidence is high that a conservative POD is
797 used, basic assumptions about toxicokinetic and toxicodynamics differences can be applied
798 as per a standard risk assessment approach. Safety assessors and regulators are used to
799 dealing with uncertainty in toxicokinetic and toxicodynamics in traditional risk assessment
800 and there are agreed frameworks to review the quality and confidence in using toxicological
801 data (SCCS, 2018). When high quality animal data are used to derive a POD for human
802 safety assessment, and using external dose metrics, an uncertainty factor of 10 (to account
803 for toxicokinetic differences) and another 10 (to account for toxicodynamic differences)
804 resulting in a margin of safety (MOS) of 100, is considered acceptable.

805 In the SCCS 10th Notes of Guidance (2018), it is explained how a POD_{sys} is calculated
806 usually from an oral toxicology study, and in the absence of any oral absorption data, the
807 oral intake is divided by 2, as it is assumed as a default that only 50% of the orally ingested
808 substance is absorbed via the gut. To calculate a systemic exposure dose (SED) following
809 dermal exposure, in the absence of valid absorption data, the starting default assumption is
810 that 50% of the dermally applied dose is absorbed.

811 Therefore, at the end of Tier 0 it is possible to achieve an acceptable MOS. If not, for
812 cosmetic ingredients, a dermal absorption value can be generated from *in vitro* human skin
813 experiments in Tier 1, and it is possible to refine the SED using this further information for
814 input to the MOS calculation. Further ADME and NAM data can also increase the confidence
815 in the POD if these new data substantiate biological similarity of the target and analogue. For
816 example, there may be toxicokinetic arguments in relation to relative potency that can be
817 made, and allow for adjustment of a POD if the target is expected to have lower or higher
818 potency than the analogues.

With these refinements, the MOS may be considered as acceptable at the end of Tier 1, given an improved level of confidence in the RAX POD and/or lower refined exposure.

If at the end of Tier 1 the MOS remains unacceptable, it is possible to refine exposure using a PBK model to estimate blood concentrations following exposures to the respective substance in experimental animals and humans. A MoIE differs from a traditional margin of exposure (MoE) in that it is calculated as the ratio of a measure of internal exposure, such as blood concentration or target-tissue dose, rather than a measure of external exposure concentration or ingested dose (Bessems et al., 2017). The ability to rely on a measure of internal rather than external exposure reduces the uncertainty in the risk assessment by incorporating chemical-specific information on the uptake, distribution, metabolism and excretion of the chemical in both the experimental animal and the human (Clewell et al. 2008). In particular, calculation of internal exposures with a PBK model can be used to replace the default uncertainty factor of 4 for interspecies differences in toxicokinetic differences (WHO, 2010). The USEPA follows this practice in determining Reference Concentrations and Reference Doses (USEPA 1994, 2006, 2011). Thus a MoIE of 25 would be equivalent to the default MOS of 100, but with greater precision for the chemical of concern. Internal exposures may also be scaled for potency of effect using a RPF derived from in vitro testing. In the caffeine (Bury et al., 2021) and parabens (Ouedraogo et al., 2021) case studies, the target substance was deemed to be a weaker antagonist/agonist than the respective analogues. Such insights can serve to increase confidence that the POD used for the risk assessment is conservative for the target or as was the case in both of our accompanying case studies the internal exposures were scaled by their respective potency factors. In this way, further refinement of uncertainty factors relating to toxicodynamic differences may be possible.

Step 10 Assessing the level of confidence for establishing if the MOS is acceptable

In order for a safety assessor to accept a RAX prediction for a cosmetics safety assessment, the assessor needs to have confidence in the accuracy and scientific credibility of the prediction and that the risk assessment where the prediction is used, is suitably conservative so as to ensure consumers are protected at the proposed exposure/use levels. This is mentioned as the last step, as this is how it may be reported best. However, in reality, one is considering the level of confidence and quality of data all the way through the process. It may be that at the end of the process, the confidence is so low in the process, that one has to re-loop to the beginning and consider alternative hypotheses. However, the process will have been transparent and scientifically rigorous in explaining whether confidence is high or low at the end of the process.

Confidence can be attained by providing sufficient high quality data, evidence and scientific rationale in the RAX documentation, and assessing against a set of defined questions, such that the predictions can be reproduced if necessary, and a full scientific critique can be performed by a body such as the SCCS in Europe. To meet these requirements for transparency and reproducibility, clear documentation, descriptions of searches and results, the databases used and on what date searches were made, etc and reporting following a consistent template structure or agreed framework would be extremely helpful.

Examples of how one could document a RAX and address levels of confidence in the evidence have begun to be discussed in Blackburn & Stuard (2014), Schultz et al (2015, 2019) and Escher et al (2019), but it is acknowledged that to support a cosmetic safety assessment, each case will have its own unique weight of evidence and levels of confidence to be considered and reported. For the purposes of hazard identification, grouping and data gap filling in the context of EU REACH, the read-across assessment framework (RAAF) was published by ECHA in 2017, which aims to codify a systematic approach for read across. Aspects of this framework for the chemistry-based parameters could be useful as applied to RAX in this framework for cosmetics. However, as each case is expected to be different in

terms of the tiers and steps needed, particularly when we are talking typically about low toxicity substances in cosmetics, more flexibility is required for a cosmetics safety assessment based on an overall risk evaluation. A specific framework such as that proposed here is needed for cosmetics safety assessment that allows for flexibility with transparency.

We propose that for a cosmetics safety assessment, a RAX justification can best be provided by following the proposed 10-step approach and by considering in Step 10 some general questions around level of confidence as described in Table 2. If a RAX justification is needed for a data gap as highlighted in a cosmetics safety dossier for a target substance, we propose a structured RAX Annex document (based upon describing the steps in the framework) is submitted together with the main safety dossier, to provide full scientific justification for a RAX. The case studies for caffeine (Bury et al., 2021) and parabens (Ouedraogo et al., 2021) are examples of how a RAX Annex document may theoretically be written up to enable scientific scrutiny and account for uncertainties in an explicit way. However, it should be noted that in 'real-life' RAX is not required for assuring the continued safety of these substances, they are used purely as exemplars. Both case studies have taken up the challenge laid out by Schultz & Cronin (2017) who having reviewed a number of read-across case studies, looking particularly at the many new types of uncertainty that arise in justifying a RAX scenario state 'Similarity in chemistry is often not enough to justify fully a read-across prediction, thus, for chronic health endpoints, toxicokinetic and/or toxicodynamic similarity is essential.' This would suggest that confidence in a POD can almost always be increased through the use of NAM, unless exposure is very low, which may be the case for some cosmetics ingredients. Therefore, it is hoped that NAM can reduce the uncertainties in RAX by generating more information on toxicodynamic, toxicokinetic, including metabolic parameters, and refining exposure estimates as demonstrated in our accompanying case studies.

Suspected analogy between chemicals e.g. in regulatory contexts, that may look alike chemically and structurally, without such scientific justification toxicologically and biologically may lead to inappropriate regulatory action.

We recognise that performing a RAX approach could in principle include a large amount of data/information in electronic format as raw data, and all search information and modelling would need to be shared to enable scientific scrutiny and reproducibility. New systems for data sharing and review and training with respect to the new data types being used may be needed for the regulator to reproduce any findings.

Bringing the analogue POD and exposure together in the preceding step, is essentially the same MOS calculation as would be covered in a standard cosmetics safety dossier today with traditional data. The level of confidence in the overall risk assessment covers potentially two separate aspects and is customised for each substance: i) the confidence that the POD for the analogue is suitably conservative to be used for the target and ii) the confidence that an exposure estimate is conservative. Consideration of both of these areas will determine the overall acceptability of the MOS.

When one has justified the level of confidence in the analogue POD, one can determine if the MOS (using external dose) or MoIE (Margin of internal exposure) would be acceptable with or without application of further uncertainty factors. Blackburn & Stuard (2014) proposed additional uncertainty factors for inclusion in the risk assessment to account for the level of confidence in the POD when using a SAR-based read across. They exemplified the approach of assessing level of confidence using case studies. It is to be underlined here that each cosmetic safety assessment has its own unique weight of evidence and levels of confidence (low, medium or high) to be considered can be informed by questions outlined in Step 10 of Table 2.

The level of confidence should also consider the degree of conservatism vs realism in the accompanying exposure estimation, and especially in the context of the exposure estimate

924 as input data into a PBK model. A tiered approach to exposure estimation is taken as
925 described earlier; from a worst case deterministic value to a realistic estimate from
926 probabilistic modelling. The positioning of PBK modelling in the context of external exposure
927 estimation as input dose information is illustrated in Figure 4.

928 *[Insert Figure 4]*

929

930 In Tier 1 of an exposure assessment (on the left hand side of Figure 4), only limited
931 information is available on the maximum % use level of a cosmetic ingredient in a product.
932 Using information in the SCCS Notes of Guidance (2018), one can then calculate a standard
933 exposure estimate in mg/kg/day. Tier 2 brings more information from survey % use levels of
934 the ingredient in products and habits and practices data. Tier 3 of an exposure assessment
935 refines the external dose exposure assessment even further by using habits and practices of
936 product use information, together with market occurrence data of the ingredient in a product
937 and specific population characteristics. These established tiers of exposure assessment are
938 different from the Tiers 0 to 2 used here in the 10-step RAX framework.

939 Cosmetics exposure assessment can be done on a single product type, or exposure can be
940 calculated for aggregate exposure scenarios, and there is specific guidance on this from the
941 SCCS (2018). In all of these cases, an external dermal dose as a mg/kg/day can be
942 calculated and this can then act as input data for a PBK model, that incorporates the aspects
943 of dermal delivery and systemic metabolism and clearance to estimate an internal metric (an
944 area under the curve; or a maximal concentration C_{max} value). Data can also be used from
945 human biomonitoring information in Tier 4 internal dose assessment, however such data can
946 be misleading if not supported by appropriate source exposure evidence for the substance
947 and PBK modelling interpretations of substance kinetics and metabolism in the human body.
948 In performing an MOS calculation using internal dose metrics, a like-for-like comparison

must be made i.e. the POD expressed as a Cmax in the blood of an animal in the toxicology study would be compared with a human exposure estimate as a Cmax in blood value.

It is important at the end of the risk assessment in Step 10, whichever method (external or internal dose metrics) is used, to discuss the scale of conservatism that is represented by the exposure estimate used in the MOS calculation. This should take account of method reliability, data quality and extent of data and reflect what, if any, additional information is required to increase confidence or whether additional uncertainty factors required.

Consideration of the use of 10-step RAX for regulatory review of cosmetic ingredients

As we have mentioned throughout, accompanying this paper, which describes the generic 10-step RAX framework, we also provide two case studies to illustrate the workflow for collating and using the data to underpin an analogue RAX-derived POD and the use of PBK modelling to derive internal exposure concentration estimates: one case study is where caffeine is the hypothetical target chemical (Bury et al., 2021) and the second where propyl paraben is the hypothetical target (Ouedraogo et al., 2021). Working through the framework for these two case studies has enabled us to demonstrate practically how a RAX supported by NAM can be used for assuring the safety of cosmetic ingredients in a regulatory context. We have built on the established concepts of RAX as based upon the principles of chemical similarity, and incorporated NAM data (*in vitro* assays, in silico profiling, PBK modelling etc) to inform the toxicokinetic and toxicodynamic aspects of an exposure driven, mode of action based RAX. The structured, step-wise approach to gather the necessary evidence for using a RAX-based approach to NGRA, is summarised in Table 2 and is intended to initiate a dialogue regarding the general inclusion of RAX-based risk assessment in future regulatory guidance for cosmetics safety assessment.

974

975

976

977 **Table 2. A 10-step RAX Framework - Aspects considered useful for regulatory review**

Aspect	Description
Tier 0	
Step 1: Identify exposure/use scenarios	• Exposure estimates based on a tiered approach, starting with a rough deterministic estimation of exposure at the low tier and evolving to a more complex subject-orientated probabilistic approach at higher tiers (e.g. as per SCCS Notes of Guidance 2018)
Step 2: Identify molecular structure of target chemical	• Review full composition of raw material, including any potential low-level impurities • Summarise structural features of the target chemical by identifying molecular scaffolds or substructures, substituents, functional groups, tautomeric forms, and alkyl chain lengths. • Identify known or likely biotransformations for the target chemical, identifying any potential for reactive metabolite formation.
Step 3: Collate supporting data on target chemical(s) and define data gap	• Collate Physico-chemical data (e.g. as per SCCS Notes of Guidance 2018) • Collate Toxicological, ADME and existing NAM information • Present all available evidence in a clear and readable tabular form • **Definition of problem formulation and the identified data gap**
Step 4: a) Identify analogue(s), b) collate existing data and c) determine similarity hypothesis (at this stage still mostly structure-based but can be refined to d) biological similarity – MOA where data is existing)	• Using structural features from steps 2 & 3, describe the analogue search strategy based on; (sub)structures, metabolism, reactive chemistries, physico-chemical properties • Define chemical (e.g. Tanimoto) similarity comparisons of molecular fingerprints plus searching by substructure with required structural features • Capture the broad search results for analogues • Collate physico-chemical data for structurally similar analogues • Collate existing toxicological and ADME data for structurally similar analogues • Collate existing NAM information for structurally similar analogues • The data should be expressed in readable tabular form, including an analysis of data quality such that it can be reviewed by a competent scientist/risk assessor • Include an analysis of concordance of data between target and analogue(s); o Can go back to start of Step 4 to broaden or narrow analogue selection/revisit analogue search strategy based on outcome of analysis of existing data set. **Derive a hypothesis of mechanism of action for target and analogues in terms of the effect to be assessed by RAX** (for example; low/no toxicity; parent versus metabolite-mediated toxicity)

	➔ Analogue(s) with good quality data and hypothesis for RAX – Go to Step 8 ➔ Analogues but insufficient data to be confident – Progress to Tier 1 N.B. If no analogues or there are analogues but no data are identified, RAX is not possible and another safety assessment strategy should be considered.
Tier 1 refinement	**Describe rationale for generating additional ADME and MoA information**
Step 5: Supporting Similar bioavailability/ADME of target chemical and analogues	• Types of data to inform on similar ADME properties – Examples include rate and extent of skin and gut permeability; extent of plasma protein binding; nature of major clearance route (metabolism or renal); rate and extent of skin, liver, plasma metabolism; likelihood of transporter effects etc
Step 6: Supporting Similar Mode/Mechanism of Action (MoA) hypothesis	• Types of data to inform on similar MoA - Untargeted gene expression or protein activity; targeted receptor/enzyme activity or cellular responses etc • The data should be described in accordance with guideline/non-guideline study requirements and expressed in clear and readable tabular form, including an analysis of data quality such that it can be reviewed by a competent scientist/risk assessor • Include an analysis of concordance of data between target and analogue(s). • Assess weight of evidence to support or refine the biological similarity hypothesis with regards to ADME and MoA. This is likely to be qualitative at this stage, serving to increase confidence in the analogue choice but could include insights into quantitative aspects that can be refined at Tier 2. ➔ Analogue(s) with good quality data and hypothesis for RAX – Go to Step 8 ➔ Analogues but insufficient data to be confident – Progress to Tier 2
Tier 2 refinement	**Describe rationale for generating more targeted/quantitative information**
Step 7: a) Targeted testing to strengthen MoA hypotheses AND/OR	• Explain what MoA is appropriate to follow up on for the safety assessment based on similarity hypothesis and toxicological relevance • A value related to MoA (e.g. Ki/IC50/AC10) should be justified and considered in relation to the internal exposure (derived in Step 7b) • The data should be described in accordance with guideline/non-guideline study requirements and expressed in clear and readable tabular form, including an analysis of data quality such that it can be reviewed by a competent scientist/risk assessor • Assess relative biological potency of target versus analogue(s) to derive a relative potency factors (RPF), as appropriate • Refine as necessary based on *in vitro* biokinetic considerations (measured or modelled)

b) Biokinetic Refinements of target chemical and analogues	• Explain what internal exposure metric is appropriate for the safety assessment based on mechanistic considerations e.g. Cmax, AUC; free or total concentration; intracellular or extracellular concentration • Search for existing PBK data (animal and human exposure); build a PBK model relevant for the target and relevant analogues • Type of data needed to parameterise PBK model – Rate and extent of skin and gut permeability; extent of plasma protein binding; rate and extent of skin, liver, plasma metabolism, clearance etc • Produce kinetic profiles for analogue (toxicity data) and human use scenarios (defined/refined in Step 1) to derive internal exposure values (defined at start of Step 7b) • Document kinetic modelling according to current best practices (WHO, 2010) • Verify model outputs (for example, using *in vivo*/human data or PK analogues). Run sensitivity and uncertainty analysis versus established criteria for PBK models ○ N.B. a degree of inaccuracy in the simulations may be acceptable if the direction of the error is described and its impact considered relative to the protection goal ➔ Analogue(s) with good quality data and hypothesis for RAX – Go to Step 8 ➔ Analogues but insufficient data to be confident – End No RAX possible
Perform Next Generation Risk Assessment	
Step 8: Performing a RAX to derive a POD	• With suitable *in chemico*, toxicology data and NAM data for the analogue(s) from either Tier 0, Tier 1 or Tier 2 evidence, and with a substantiated hypothesis of similarity between target and source chemicals, the POD from the most suitable analogue(s) can be used as a basis for deriving a POD for the target chemical.

Step 9: Perform a MOS/MoIE evaluation using RAX	• Tier 0 MOS ≥100: POD source / target exposure
	• Tier 1 MOS ≥100: POD source / target exposure
	  ○ 100 accounts for toxicokinetic and toxicodynamic differences between species and between individuals. There may be a need to allow for some adjustment of the acceptable MOS by using systemic exposure dose and kinetic assumptions e.g. default oral absorption and consideration of data uncertainties etc
	• Tier 2 MoIE ≥25: POD_{sys} source / (target $exposure_{sys}$ X RPF^{*})
	  ○ Assumes that kinetics are fully accounted for which allows the interspecies TK UF to be set to 1
	  ▪ *if NAM data allow
	  ○ Further adjustment of UFs up or down may be needed depending upon uncertainties in the data and using toxicodynamic information
	
	• Go to Step 10 to assess confidence in the risk assessment

Step 10: Assessing confidence in the risk assessment	• Describe overall level of confidence (low, medium or high) that the RAX is appropriately conservative as part of an exposure driven risk assessment. Throughout the whole process, uncertainties and the level of confidence in the data should be captured as transparently as possible and integrated to provide and overall level of confidence in the assessment. Examples of questions to address: ○ What type of category formation was attempted and was it suitable for the context of the read-across? ○ How well made was the premise or hypothesis of the read-across argument? ○ What rationale was used to select the NAMs used and how did they support the decision making? ○ How was mechanism of action considered supported and assessed? ○ How was similarity in TD/effects defined and assessed to support the MoA? ○ How was similarity in TK/potency defined and assessed? ○ What were the uncertainties in the toxicological data for read-across data and how did they allow for an assessment of robustness of these data? ○ How were NAMs applied and did they assist in the reduction of uncertainty? ○ What are the key strengths of the case study? ○ What are the key limitations of the case study? • If overall confidence is not acceptable e.g. method reliability, data quality and extent of data, what additional information is required to increase confidence or are additional uncertainty factors required?

978

979

Discussion

In this paper, we have built on the EU SEURAT-1 programme and ICCR concepts and further developed a practical and structured 10-step framework to illustrate how RAX using NAM can contribute to consumer safety assessment for cosmetic ingredients. We propose a Tiered exposure-driven and evidence-based approach to RAX, where NAMs are used to strengthen a mode/mechanism of action hypothesis and support data on kinetics and exposure. Accompanying this paper we have presented two case studies (Bury et al., 2021; Ouedraogo et al., 2021) which will help to illustrate how this framework can be followed in practice. Further case studies and full RAX submissions in the form of regulatory dossiers will help to cement how this framework can be followed in practice, determine which kinds of data are most helpful for supporting RAX and over time inform on what the magnitude of the MOS/MoIE should be based on a wider experience to assure safety.

RAX becomes more customised as one progresses through the Tiers of the Framework, e.g. it differs depending on the problem formulation, MoA information, data availability, resources available to support further experimentation or data gathering. Even in absence of formal 'validation, it can be seen that scientifically valid NAM approaches can be useful for risk assessment. The challenge comes in how to document and report the approaches taken in a credible and reproducible way such that a regulatory scientist can review in a step-wise way and critique with transparency. The possibilities for using different types of evidence depending on the problem formulation, leads to an approach based on good science but one that is by nature less prescriptive and more flexible.

Using toxicology data on similar substances for risk assessment, carries a level of uncertainty but risk assessors are used to dealing with uncertainties through application of uncertainty factors, usually around toxicokinetic and toxicodynamic differences. The level of confidence analysis in a RAX approach based on chemical similarity alone is qualitative and somewhat relying on expert judgements and narrative statements. The most straight-forward

case of NAM is to underpin and increase confidence in a POD for a target chemical, reading across from toxicology data generated for a similar analogue(s), and one can exit the framework when one is sufficiently confident in the POD selected. There has been an expansion in the number of tools and databases in the past decade, including for physico-chemical and ADME analyses, *in vitro* data and toxicogenomics data etc, that can be used to underpin analogue(s) identification in terms of similarity (Pawar et al., 2019), many of which we used in our parabens case study (Ouedraogo et al., 2021). It is therefore difficult to see that a single unified approach to searching, analysing and reporting out of these new tools could be done, as it would limit possibilities of using the best scientific tools of the day. Therefore, a general approach for analogue selection based on sound scientific principles that can be supported with a variety of tools and databases could be used so long as they are explained well and justified. The scientific method for analogue selection should be reported using a structured narrative, and description of the adopted search process and tools provided in such detail that the process and conclusions drawn can be reproduced by a regulator and reviewed by a competent and trained professional risk assessor with experience of the tools being used in the submitted dossier. It is important that close working relationships exist as new tools develop, between the regulatory scientists, academia and industry, so that working knowledge and confidence of NAMs grows in all sectors relevant to performing safety assessment. It is envisaged that as knowledge on adverse outcome pathways (AOPs) develops, for defined modes and mechanisms of action, standard batteries of NAMs will make it easier to support a targeted rationale but equally important is the ability to cast a wide net and cover a broad biological space with data such as toxicogenomics, to ensure no unexpected modes of action for a target substance compared to its analogue(s). As we have seen in both of our case studies, targeted toxicodynamic differences, e.g. differences in the binding affinity at a target receptor, can also result in relative potency differences between target and analogue that could in principle be considered in the risk assessment. However, one must be confident that the measure on which the RPF is based, is the driver for a potential potency difference in humans.

Some of the biggest uncertainties in RAX have been attributed to metabolism and toxicokinetics between species and between target and analogues substances. NAMs that are used to generate data on ADME properties are useful to compare substance behaviours. Metabolite analysis in cells and tissues is today relatively straight forward using mass spectroscopy and other analytical chemistry techniques. PBK modelling can generate information on internal concentrations in blood/plasma or target organs, in order to reduce uncertainty in exposure considerations in relation to in route-to-route extrapolations, between species and between target and analogue substances. Therefore, when relying on internal exposures a lower MOS is acceptable in the risk assessment since an uncertainty factor for interspecies kinetic differences is not necessary. PBK models also may be used to determine if differences in kinetics might lead to differences in effects or relative potency for the target substance vs the analogues.

It is a long term ambition of the cosmetics sector to derive an approach that relies on no animal data what so ever, which requires confidence in NAMs to cover the breadth and depth of the known world of mechanistic toxicology and how to implement the knowledge within the context of human adverse outcome pathways. RAX is seen as important stepping stone in this journey. In a report from the Regulators-Industry Joint Working Group (JWG) of the International Cooperation on Cosmetics Regulation (ICCR) the potential for use of NAMs in an *ab initio* risk assessment is outlined, where the approach is exposure-driven, NAM are used to inform a MoA hypothesis and the possibilities for performing a QIVIVE approach in deriving a POD using *in vitro* data exist (ICCR, 2017b, 2018). We see RAX as a vital element also in this discussion to support the generation of a confident MoA hypothesis, i.e. in showing that similar structures may have a common MoA, even if the data are missing to conclude the risk assessment based on read across.

In summary, our 10-step read-across (RAX) framework builds on established approaches for defining chemical similarity by substantiating hypotheses on toxicological similarity of substances using NAM in both the biological and kinetic domains. A next generation risk

assessment (NGRA) may then be performed with an acceptable level of confidence for a data-poor target substance, based on RAX approaches using data-rich analogue(s) with integration of kinetic/dynamic differences as appropriate.

Funding body information

This work was funded through the Long Range Science Strategy (LRSS) programme of Cosmetics Europe (https://www.lrsscosmeticseurope.eu).

Declaration of competing interest

No known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the helpful comments provided by Corie Ellison, Cathy Lester, Jane Rose and Sharon Stuard.

References

Ankley, G.T. & Edwards, S.W. 2018. The Adverse Outcome Pathway: A Multifaceted Framework Supporting 21st Century Toxicology. *Current Opinions in Toxicology* 9, 1-7. doi: 10.1016/j.cotox.2018.03.004

Bajusz, D., Rácz, A., Héberger, K., 2015. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Chemoinformatics* 7, 20.

Ball, N., Bartels, M., Budinsky, R., Klapacz, J., Hays, S., Kirman, C., Patlewicz, G., 2014. The challenge of using read-across within the EU REACH regulatory framework; how much uncertainty is too much? Dipropylene glycol methyl ether acetate, an exemplary case study. *Regul Toxicol and Pharmacol* 68, 212–221.

Ball, N., Cronin, M.T.D., Shen, J., Blackburn, K., Booth, E.D., Bouhifd, M., Donley, E., Egnash, L., Hastings, C., Juberg, D.R., Kleensang, A., Kleinstreuer, N., Kroese, E.D., Lee, A.C., Luechtefeld, T., Maertens, A., Marty, S., Nacif,f J.M., Palmer, J., Pamies, D., Penman, M., Richarz, A-N., Russo, D.P., Stuard, S.B., Patlewicz, G., van Ravenzwaay, B., Wu, S., Zhu, H., Hartung, T., 2016. Toward Good Read-Across Practice (GRAP) Guidance. *ALTEX-ALTERNATIVES TO ANIMAL EXPERIMENTATION* 33, 149-166. doi 10.14573/altex.1601251

Benfenati E, Chaudhry Q, Gini G, Dorne JL 2019 Integrating in silico models and read-across methods for predicting toxicity of chemicals: A step-wise strategy. Environment International, 131.

https://doi.org/10.1016/j.envint.2019.105060.

Berggren, E., Amcoff, P., Benigni, R., Blackburn, K., Carney, E., Cronin, M., Deluyker, H., Gautier, F., Judson, R.S., Kass, G.E., Keller, D., Knight, D., Lilienblum, W., Mahony, C., Rusyn, I., Schultz, T., Schwarz, M., Schüürmann, G, White, A., Burton, J., Lostia, A.M., Munn, S., Worth, A., 2015. Chemical Safety Assessment Using Read-Across: Assessing the Use of Novel Testing Methods to Strengthen the Evidence Base for Decision Making. *Environ Health Perspect.* 123(12), 1232–1240. doi: 10.1289/ehp.1409342.

Berggren, E., White, A., Ouedraogo, G., Paini, A., Richarz, A-N., Bois, F.Y., Exner, T., Leite, S., Grunsven, L.Av., Worth, A. et al. 2017. *Ab initio* chemical safety assessment: A workflow based on exposureconsiderations and non-animal methods. *Computational Toxicology*. 4, 31-44.

Bessems, J.G.M., Paini, A., Gajewska, M., Worth, A., 2017. The margin of internal exposure (MOIE) concept for dermal risk assessment based on oral toxicity data – A case study with caffeine. *Toxicology* 392, 119–129.

Blackburn, K., Stuard, S.B., 2014. A framework to facilitate consistent characterization of read across uncertainty. *Regul Toxicol Pharmacol*. 68(3), 353-362. doi:10.1016/j.yrtph.2014.01.004

Blackburn, K., Carr, G., Rose, J.L., Selman, B.G., 2020. An interim internal Threshold of Toxicologic Concern (iTTC) for chemicals in consumer products, with support from an automated assessment of ToxCast™ dose response data. *Regul Toxicol Pharmacol* 114, 104656 . doi.org/10.1016/j.yrtph.2020.104656

Bury, D., Alexander-White, C., Clewell III, H.J., Cronin, M., Desprez, B., Detroyer, A., Efremenko, A., Firman, J., Hack, E., Hewitt, N.J., Kenna, G., Klaric, M., Lester, C., Mahony, C., Ouedraogo, G., Paini, A., Schepky, A., 2021. New Framework for a Non-Animal Approach Adequately Assures the Safety of Cosmetic Ingredients – A Case Study on Caffeine. *Regul Toxicol Pharmacol* Apr 24;123:104931. doi: 10.1016/j.yrtph.2021.104931. Online ahead of print.

Campbell, J.L., Yoon, M., Clewell, H.J., 2015). A case study on quantitative *in vitro* to *in vivo* extrapolation for environmental esters: Methyl-, propyl- and butylparaben. *Toxicology* 2015, 332, 67-76. doi:10.1016/j.tox.2015.03.010.

Cereto-Massagué, A., Ojeda, M.J., Valls, C., Mulero, M., Garcia-Vallvé, S., Pujadas, G., 2015. Molecular fingerprint similarity search in virtual screening. *Methods* 71, 58–63.

Chen, H., Chidboy, M.A., Robinson, J.F., 2020. Retinoids and developmental neurotoxicity: Utilizing toxicogenomics to enhance adverse outcome pathways and testing strategies. *Reprod Toxicol.* 96, 102-113. doi:10.1016/j.reprotox.2020.06.007

Clewell III, H.J., Andersen, H.J., and Blaauboer, B.J. 2008. On the incorporation of chemical-specific information in risk assessment. *Toxicology Letters* 180, 100–109.

Coulet, M., Latado, H., Moser, M., Besselink, H., Tate, M., Minetto, F., Fontannaz, C.C., Serrant, P., Mollergues, J., Piguet, D., Schilter, B., Marin-Kuan, M.,2019. Use of *in vitro* bioassays to facilitate read-across assessment of nitrogen substituted heterocycle analogues of polycyclic aromatic hydrocarbons. *Toxicol In Vitro* 59, 281-291. doi:10.1016/j.tiv.2019.04.030 2019

Cronin, M.T.D., Madden, J.C., Yang, C., Worth, A.P., 2019. Unlocking the potential of in silico chemical safety assessment - A report on a cross-sector symposium on current opportunities and future challenges. *Comput Toxicol*. 10, 38–43. doi:10.1016/j.comtox.2018.12.006

Cronin, M.T.D., Enoch, S.J., Mellor, C.L., Przybylak, K.R., Richarz, A.N., Madden, J.C., 2017. *In Silico* Prediction of Organ Level Toxicity: Linking Chemistry to Adverse Effects. *Toxicol Res*. 33(3), 173-182. doi:10.5487/TR.2017.33.3.173

Cronin, M.T.D., Richarz, A.N.,2017. Relationship between Adverse Outcome Pathways and chemistry-cased *in silico* models to predict toxicity. *Applied in vitro Toxicology* 3: 286-297. DOI: 10.1089/aivt.2017.0021

De Abrew, K.N., Overmann, G.J., Adams, R.L., Tiesman, J.P., Dunavent, J., Shan, Y.K., Carr, G.J., Daston, G.P., Naciff, J.M., 2015. A novel transcriptomics based *in vitro* method to compare and predict hepatotoxicity based on mode of action. *Toxicology* 328, 29-39. doi:10.1016/j.tox.2014.11.008.

De Abrew, K.N., Shan, Y.K., Wang, X., Krailler, J.M., Kainkaryam, R.M., Lester, C.C., Settivari, R.S., LeBaron, M.J., Naciff, J.M., Daston, G.P., 2019. Use of connectivity mapping to support read across: A deeper dive using data from 186 chemicals, 19 cell lines and 2 case studies. *Toxicology* 423, 84-94. doi: 10.1016/j.tox.2019.05.008. Epub 2019 May 21.

Delmaar, J.E. and van Engelen, J.G.M., 2006. Aggregating Human Exposure to Chemicals and Overview of Tools and Methodologies. RIVM report 630700001/2006

Dent, M., Teixeira Amaral, R., Amores Da Silva, P., Ansell, J., Boisleve, F., Hatao, M., Hirose, A., Kasai, Y., Kern, P., Kreiling, R., Milstein, S., Montemayor, B., Oliveira, J., Richarz, A., Taalman, R., Vaillancourt, E., Verma, V., O'Reilly Cabral Posada, N.V., Wiess, C., Kojima, H., 2018. Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. Computational Toxicology 7, 20-26. doi: 10.1016/j.comtox.2018.06.001

ECHA 2016. New Approach Methodologies in Regulatory Science Proceedings of a scientific workshop **ISBN:** 978-92-9495-397-1 **DOI:** 10.2823/543644.

ECHA 2017. Read-across assessment framework (RAAF) ECHA-17-R-01-EN, ISBN 978-92-9495-758-0. https ://doi.org/10.2823/61921 2

EFSA 2019. Guidance on the use of the Threshold of Toxicological Concern approach in food safety assessment. EFSA J 17(6) . doi: 10.2903/j.efsa.2019.5708

Ellison, C. A., Blackburn, K. L., Carmichael, P. L., Clewell III, H. J., Cronin, M., 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. *Regulatory toxicology and pharmacology : RTP*, *103*, 63–72. https://doi.org/10.1016/j.yrtph.2019.01.016

Ellison, C.A. & Wu, S. 2020. Application of structural and functional pharmacokinetic analogs for physiologically based pharmacokinetic model development and evaluation. Reg Tox Pharm 114, 104667. doi.org/10.1016/j.yrtph.2020.104667

European Medicines Agency (EMA) 2016. Guideline on the qualification and reporting of 5 physiologically based pharmacokinetic (PBPK) modelling 6 and simulation. EMA/CHMP/458101/2016. July 2016. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-qualification-reporting-physiologically-based-pharmacokinetic-pbpk-modelling_en.pdf

Embry, M.R., Bachman, A.N., Bell, D.R., Boobis, A.R., Cohen, S.M., Dellarco, M., Dewhurst, I.C., Doerrer, N.G., Hines, R.N., Moretto, A., Pastoor, T.P., Phillips, R.D., Rowlands, J.C., Tanir, J.Y., Wolf, D.C., Doe, J.E. 2014. Risk assessment in the 21st century: Roadmap and matrix, *Critical Reviews in Toxicology*, 44:sup3, 6-16, DOI: 10.3109/10408444.2014.931924.

Escher, S.E., Kamp, H., Bennekou, S.H., Bitsch, A., Fisher, C., Graepel, R., Hengstler, J.G., Herzler, M., Knight, D., Leist ,M., Norinder, U., Ouédraogo, G., Pastor, M., Stuard, S., White, A., Zdrazil, B., van de Water, B., Kroese, D., 2019. Towards grouping concepts based on new approach methodologies in chemical hazard assessment: the read-across approach of the EU-ToxRisk project *Arch Toxicol.* 93(12), 3643-3667. doi: 10.1007/s00204-019-02591-7.

Firman, J.W., Patel, A., Date, M., Cronin, M.T.D., Schultz, T.W. 2018. Read-across of 90-day rodent repeated-dose toxicity: A case study for selected simple aryl alcohol alkyl carboxylic acid esters. *Computational Toxicology* 7: 1-8. DOI: 10.1016/j.comtox.2018.05.001

Fisher, J.W., Twaddle, N.C., Vanlandingham, M., Doerge, D.R., 2011. Pharmacokinetic modeling: Prediction and evaluation of route dependent dosimetry of bisphenol A in monkeys with extrapolation to humans. *Toxicol Appl Pharmacol* 257, 122–136

Gelbke, H.P., Ellis-Hutchings, R., Muellerschoen, H., Murphy, S., Pemberton, M., 2018. Toxicological assessment of lower alkyl methacrylate esters by a category approach. *Regul Toxicol Pharmacol* 92, 104-127 10.1016/j.yrtph.2017.11.013

Groothuis, F.A., Heringa, M.B., Nicol, B., Hermens, J.L., Blaauboer, .BJ., Kramer, N.I., 2015. Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. *Toxicology*. 332:30-40. doi:10.1016/j.tox.2013.08.012. Epub 2013 Aug 23. PMID: 23978460.

Hanway, R.H., Evans, P.F., 2000. Read-across of toxicological data in the notification of new chemicals. *Toxicology Letters* 116 (Suppl. 1): 61

Helman, G., Shah, I., Patlewicz, G., 2018. Extending the Generalised Read-Across approach (GenRA): A systematic analysis of the impact of physicochemical property information on read-across performance. *Comput Toxicol.* 2018;8:34-50. doi: 10.1016/j.comtox.2018.07.001.

Helman, G., Shah, I., Williams, A.J., Edwards, J., Dunne, J., Patlewicz, G. 2019. Generalized Read-Across (GenRA): A Workflow Implemented into the EPA CompTox Chemicals Dashboard ALTEX-ALTERNATIVES TO ANIMAL EXPERIMENTATION 36, 462-465 10.14573/altex.1811292 2019

Hewitt, N.J., Grégoire, S., Cubberley, R., Duplan, H., Eilstein, J., Ellison, C., Lester, C., Fabian, E., Fernandez, J., Géniès, C., Jacques-Jamin, C., Klaric, M., Rothe, H., Sorrell, I., Lange, D., Schepky, A. 2020. Measurement of the penetration of 56 cosmetic relevant chemicals into and through human skin using a standardised protocol. *J Appl Toxicol.* 40(3):403-415. doi: 10.1002/jat.3913.

House, J. S., Grimm, F. A., Klaren, W. D., Dalzell, A., Kuchi, S., Zhang, S.-D., Lenz, K., Boogaard, P. J., Ketelslegers, H. B., Gant, T. W., Wright, F. A. and Rusyn, I. (2021) “Grouping of UVCB substances with new approach methodologies (NAMs) data”, ALTEX - Alternatives to animal experimentation, 38(1), pp. 123-137. doi: 10.14573/altex.2006262.International Cooperation on Cosmetics Regulation (ICCR) (2017) ‘Integrated Strategies for Safety Assessment of Cosmetic Ingredients – Part I’. https://iccr-cosmetics.org/topics/

International Cooperation on Cosmetics Regulation (ICCR) (2018) ‘Integrated Strategies for Safety Assessment of Cosmetic Ingredients – Part II’. https://iccr-cosmetics.org/topics/

Jiang, J., Pieterman, C.D., Ertaylan, G., Peeters, R.L.M., de Kok T.M.C.M., 2019. The application of omics-based human liver platforms for investigating the mechanism of drug-induced hepatotoxicity *in vitro*. *Arch Toxicol*. 2019;93(11):3067-3098. doi:10.1007/s00204-019-02585-5.

Klimisch, H.J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. 25(1),1-5.

1235 Kovarich, S., Ceriani, L., Gatnik, M.F., Bassanc, A., Pavan, M., 2019. Filling Data Gaps by Read-
1236 across: A Mini Review on its Application, Developments and Challenges MOLECULAR
1237 INFORMATICS 38. 10.1002/minf.201800121 2019

1238 Krewski, D., Andersen, M.E., Tyshenko, M.G., Krishnan, K., Hartung, T., Boekelheide, K., Wambaugh
1239 J.F., Jones, D., Whelan, M., Thomas, R., Yauk, C., Barton-Maclaren, T., Cote, I.2020. Toxicity testing
1240 in the 21st century: progress in the past decade and future perspectives. *Arch Toxicol.* 94(1):1-58.
1241 doi: 10.1007/s00204-019-02613-4. Epub 2019 Dec 17.

1242 Laroche, C., Aggarwal, M., Bender, H., Benndorf, P., Birk, B., Crozier, J., Dal Negro, G., De Gaetano,
1243 F., Desaintes, C., Gardner, I., Hubesch, B., Irizar, A., John, D., Kumar, V., Lostia, A., Manou, I.,
1244 Monshouwer, M., Mueller, B.P., Paini, A., Reid, K., Rowan, T., Sachana, M., Schutte, K., Stirling, C.,
1245 Taalman, R., van Aert, L., Weissenhorn, R., Sauer Ursula, G. 2018. Finding synergies for 3Rs-
1246 Toxicokinetics and read-across: Report from an EPAA partners' Forum *Regul Toxicol Pharmacol* 99,
1247 5-21 doi 10.1016/j.yrtph.2018.08.006

1248 Lester, C., Reis, A., Laufersweiler, M., Wu, S., Blackburn, K. 2018. Structure activity relationship
1249 (SAR) toxicological assessments: The role of expert judgment. *Regul Toxicol Pharmacol* 92, 390-406.

1250 Lipinski, C.A., Lombardo, F., Dominy, B.W., & Feeney, P.J., 2001. Experimental and computational
1251 approaches to estimate solubility and permeability in drug discovery and development settings, *Adv.*
1252 *Drug Deliv. Rev*. 46(2001), 3–26.
1253
1254 Little, J.G., Marsman, D.S., Baker, T.R., Mahony, C., 2017. In silico approach to safety of botanical
1255 dietary supplement ingredients utilizing constituent-level characterization. *Food Chem Toxicol*. 107(Pt
1256 A):418-429. doi:10.1016/j.fct.2017.07.017.
1257
1258 Liu, Z., Huang, R., Roberts, R., Tong, W., 2019. Toxicogenomics: A 2020 Vision. *Trends Pharmacol*
1259 *Sci* 40(2):92-103. doi: 10.1016/j.tips.2018.12.001.

1260 Low, Y., Sedykh, A., Fourches, D., Golbraikh A., Whelan, M., Rusyn, I., Tropsha, A., 2013. Integrative
1261 Chemical–Biological Read-Across approach for chemical hazard classification. *Chemical Research in*
1262 *Toxicology* 26: 1199-1208. doi: 10.1021/tx400110f

1263 Luijten, M., Ball, N.S., Dearfield, K.L., Gollapudi, B.B., Johnson, G.E., Madia, F., Peel, L., Pfuhler, S.,
1264 Settivari, R.S., Ten Burg, W., White, P.A., van Benthem, J., 2020. Utility of a next generation
1265 framework for assessment of genomic damage: A case study using the industrial chemical benzene.
1266 *Environ Mol Mutagen*. 61(1), 94-113. doi: 10.1002/em.22346.

1267 Madden, J.C., Pawar, G., Cronin, M.T.D., Webb, S., Tan, Y-M., Paini, A., 2019. *In silico* resources to
1268 assist in the development and evaluation of physiologically-based kinetic models. *Computational*
1269 *Toxicology* 11, 33-49. DOI: 10.1016/j.comtox.2019.03.001

1270 McMullin, T. S., Yang, Y., Campbell, J., Clewell, H. J., Plotzke, K., & Andersen, M. E., 2016.
1271 Development of an integrated multi-species and multi-dose route PBPK model for volatile methyl
1272 siloxanes - D4 and D5. *Regul toxicol and pharmacol 74 Suppl*, S1–S13.

1273 Meek, M.E., Boobis, A., Crofton, K., Heinemeyer, G., Van Raaij, M., Vickers, C., 2011. Risk
1274 Assessmet of combined exposure to multiple chemicals: a WHO/IPCS framework. *Regul. Toxicol.*
1275 *Pharmacol*. 60, S1-S14.

1276 Mellor, C.L., Steinmetz, F.P., Cronin, M.T.D., 2016a. The identification of nuclear receptors
1277 associated with hepatic steatosis to develop and extend adverse outcome pathways. *Critical Reviews*
1278 *in Toxicology* 46, 138-152. DOI: 10.3109/10408444.2015.1089471

1279 Mellor, C.L., Steinmetz, F.P., Cronin, M.T.D., 2016b. Using molecular initiating events to develop a
1280 structural alert based screening workflow for nuclear receptor ligands associated with hepatic
1281 steatosis. *Chem Res Toxicol* 29, 203-212. DOI: 10.1021/acs.chemrestox.5b00480

1282 Mellor, C.L., Schultz, T.W., Przybylak, K.R., Richarz, A-N., Bradbury, S.P., Cronin, M.T.D., 2017.
1283 Read-across for rat oral gavage repeated-dose toxicity for short-chain mono-alkylphenols: A case
1284 study. *Comp Toxicol* 2: 1-11. DOI: dx.doi.org/10.1016/j.comtox.2017.03.003.

1285 Mellor, C.L., Marchese Robinson, R.L., Benigni, R., Ebbrell, D., Enoch, S.J., Firman, J.W., Madden,
1286 J.C., Pawar, G., Yang, C., Cronin, M.T.D., 2019. Molecular fingerprint-derived similarity measures for

toxicological read-across: Recommendations for optimal use. *Regul Toxicol Pharmacol.* 101, 121-134. doi: 10.1016/j.yrtph.2018.11.002.

Moroni, L., Barbaro, F., Caiment, F., Coleman, O., Costagliola, S., Conza, G.D., Elviri, L., Giselbrecht, S., Krause, C., Mota, C., Nazzari, M., Pennington, S.R., Ringwald, A., Sandri, M., Thomas, S., Waddington, J., Toni, R., 2020. SCREENED: A Multistage Model of Thyroid Gland Function for Screening Endocrine-Disrupting Chemicals in a Biologically Sex-Specific Manner. *Int J Mol Sci.* 21(10), 3648. Published 2020 May 21. doi:10.3390/ijms21103648.

National Academies of Sciences, Engineering, and Medicine (NAS) 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC: The National Academies Press. doi: 10.17226/24635.

OECD 2014 Guidance on the Grouping of Chemicals. 2nd edition. Document accessed on 10 June 2021 at https://www.oecd.org/publications/guidance-on-grouping-of-chemicals-second-edition-9789264274679-en.htm

OECD 2021 Guidance document on the characterisation, validation and reporting of PBK models for regulatory purposes. ENVIRONMENT DIRECTORATE CHEMICALS AND BIOTECHNOLOGY COMMITTEE monograph ENV/CBC/MONO(2021)1. 3 Feb 2021.

Ouedraogo, G., Alexander-White, C., Bury, D., Clewell III, H.J., Cronin, M., Cull T., Dent, M., Desprez, B., Detroyer, A., Ellison, C., Giammanco, S., Hack, E., Hewitt, N.J., Kenna, G., Klaric, M., Kreiling, R., Lester, C., Mahony, C., Mombelli, E., Naciff, J., O'Brien, J., Schepky, A., Tozer, S., van der Burg, B., van Vugt, B., Stuard, S., 2020. Read-Across and New Approach Methodologies applied in a 10-Step Framework to Assure the Safety of Cosmetic Ingredients – A Case Study with Parabens. Manuscript submitted to *Regul Toxicol Pharmacol*. Special Issue.

Patlewicz, G., Ball, N., Boogaard, P.J., Becker, R.A., Hubesch, B., 2015. Building scientific confidence in the development and evaluation of read-across. *Regul Toxicol Pharmacol.* 72(1), 117-33. doi: 10.1016/j.yrtph.2015.03.015.

Patlewicz, G., George, H., Prachi, P., Imran, S., 2017. Navigating through the minefield of read-across tools: A review of in silico tools for grouping. *Comput Toxicol*. 3, 1–18. doi:10.1016/j.comtox.2017.05.003

Pawar, G., Madden, J.C., Ebbrell, D., Firman, J.W., Cronin, M.T.D., 2019. *In Silico* Toxicology Data Resources to Support Read-Across and (Q)SAR. *Front Pharmacol*. 10, 561. doi: 10.3389/fphar.2019.00561. eCollection 2019.

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 β-olefinic alcohols. *Comp Toxicology* 1, 22-32. DOI: dx.doi.org/10.1016/j.comtox.2016.11.001

Punt, A., Bouwmeester, H., Blaauboer, B. J., Coecke, S., Hakkert, B., Hendriks, D. F. G., Jennings, P., Kramer, N. I., Neuhoff, S., Masereeuw, R., Paini, A., Peijnenburg, A. A. C. M., Rooseboom, M., Shuler, M. L., Sorrell, I., Spee, B., Strikwold, M., van der Meer, A. D., van der Zande, M., Vinken, M., Yang, H., Bos, P. M. J. and Heringa, M. B., 2020. New approach methodologies (NAMs) for human-relevant biokinetics predictions: Meeting the paradigm shift in toxicology towards an animal-free chemical risk assessment, *ALTEX - Alternatives to animal experimentation*, 37(4):607-622. doi:10.14573/altex.2003242.

Ryan KR, Huang MC, Ferguson SS, Waidyanatha S, Ramaiahgari S, Rice JR, Dunlap PE, Auerbach SS, Mutlu E, Cristy T, Peirfelice J, DeVito MJ, Smith-Roe SL, Rider CV. Evaluating Sufficient Similarity of Botanical Dietary Supplements: Combining Chemical and In Vitro Biological Data. Toxicol Sci. 2019 Dec 1;172(2):316-329. doi: 10.1093/toxsci/kfz189. PMID: 31504990; PMCID: PMC6876258.

Sauer, U.G., Barter, R.A., Becker, R.A., Benfenati, E., Berggren, E., Hubesch, B., Hollnagel, H.M., Inawaka, K., Keene, A.M., Mayer, P., Plotzke, K., Skoglund, R., Albert, O., 2020. 21st Century Approaches for Evaluating Exposures, Biological Activity, and Risks of Complex Substances: Workshop highlights. *Regul Toxicol Pharmacol.* 111, 104583. doi: 10.1016/j.yrtph.2020.104583.

SCCS 2018. The SCCS Notes Of Guidance For The Testing Of Cosmetic Ingredients And Their Safety Evaluation, 10th revision. SCCS/1602/18. Adopted 24-25 October 2018.

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

Schultz, T.W., Przybylak, K.R., Richarz, A-N., Mellor, C.L., Bradbury, S.P., Cronin, M.T.D., 2017b. Read-across of 90-day rat oral repeated-dose toxicity: A case study for selected 2-alkyl-1-alkanols. *Comput Toxicology* 2, 28-38. DOI: https://doi.org/10.1016/j.comtox.2017.02.005

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 Toxicology* 9: 1-11. DOI: 10.1016/j.comtox.2018.10.003

Shah, I., Liu, J., Judson, R.S., Thomas, R.S., Patlewicz, G., 2016. Systematically evaluating read-across prediction and performance using a local validity approach characterised by chemical structure and bioactivity information. *Regul Toxicol Pharmacol*. 79, 12-24. doi: 10.1016/j.yrtph.2016.05.008.

Shen, J., Kromidas, L., Schultz, T., Bhatia, S., 2014. An in silico skin absorption model for fragrance materials. Food Chem Toxicol 74, 164-176.

Solomon KR, Brock TCM, de Zwart D, Dyer SD, Posthuma L, Richards SM, Sanderson H, Sibley PK, van den Brink PJ (2008) Extrapolation in the context of criteria setting and risk assessment. In Extrapolation practice for ecotoxicological effect characterization of chemicals. Eds Brock et al SETAC Press & CRC Press, pages 1-32. ISBN 9781420073904

Sperber, S., Wahl, M., Berger, F., Kamp, H., Lemke, O., Starck, V., Walk, T., Spitzer, M., Ravenzwaay, B. V., 2019. Metabolomics as read-across tool: An example with 3-aminopropanol and 2-aminoethanol. *Regul Toxicol Pharmacol* 108, 104442. 10.1016/j.yrtph.2019.104442

Tan, Y.M., Chan, M., Chukwudebe, A., Domoradzki, J., Fisher, J., Hack, C.E., Hinderliter, P., Hirasawa, K., Leonard, J., Lumen, A., Paini, A., Qian, H., Ruiz, P., Wambaugh, J., Zhang, F., Embry, M., 2020. PBPK model reporting template for chemical risk assessment applications. *Regul Toxicol Pharmacol*. 115, 104691. doi:10.1016/j.yrtph.2020.104691

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, 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. doi: 10.1093/toxsci/kfz058.

Troutman, J.A., Rick, D.L., Stuard, S.B., Fisher, J., Bartels, M.J., 2015. Development of a Physiologically-Based Pharmacokinetic Model of 2-Phenoxyethanol and its Metabolite Phenoxyacetic Acid in Rats and Humans to Address Toxicokinetic Uncertainty in Risk Assessment. *Regul Toxicol Pharmacol*, doi: 10.1016/j.yrtph.2015.07.012

US Environmental Protection Agency (USEPA) 1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry, EPA/600/8-90/066F

US Environmental Protection Agency (USEPA) 2006. Approaches for the Application of Physiologically Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment. National Center for Environmental Assessment, Washington, DC; EPA/600/R-05/043F. Available from: National Technical Information Service, Springfield, VA, and online at http://epa.gov/ncea

US Environmental Protection Agency (USEPA) 2011. Recommended use of body weight3/4 as the default method in derivation of the oral reference dose. Risk Assessment Forum. EPA/100/R11/0001

VanderMolen, K.M., Naciff, J.M., Kennedy, K., Otto-Bruc, A., Shan, Y., Wang, X., Daston, G.P., Mahony, C., 2020. Incorporation of *In vitro* Techniques for Botanicals Dietary Supplement Safety

Assessment - Towards evaluation of Developmental and Reproductive Toxicity (DART). *Food Chem Toxicol.* 6, 111539. doi:10.1016/j.fct.2020.111539.

Van der Stel W, Carta G, Eakins J, Delp J, Suciu I, Forsby A, Cediel-Ulloa A, Attoff K, Troger F, Kamp H, Gardner I, Zdrazil B, Moné MJ, Ecker GF, Pastor M, Gómez-Tamayo JC, White A, Danen EHJ, Leist M, Walker P, Jennings P, Hougaard Bennekou S, Van de Water B. New approach methods supporting read-across: Two neurotoxicity AOP-based IATA case studies. ALTEX. 2021 Jun 10. doi: 10.14573/altex.2103051. Epub ahead of print. PMID: 34114044.

van Ravenzwaay, B., Sperber, S., Lemke, O., Fabian, E., Faulhammer, F., Kamp, H., Mellert, W., Strauss, V., Strigun, A., Peter, E., Spitzer, M., Walk, T., 2016. Metabolomics as read-across tool: A case study with phenoxy herbicides *Regul Toxicol Pharmacol* 81, 288-304. doi:10.1016/j.yrtph.2016.09.013

WHO 2010. Characterisation and Application of Physiologically Based Pharmacokinetic models in Risk Assessment. International Programme on Chemical Safety. Harmonization Project Document No. 9 ISBN 978 92 4 150090 6.

Willett, P., Barnard, J., Downs, G., 1998. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 38, 983–996.

Wu, S., Blackburn, K., Amburgey, J., Jaworska, J., Federle, T.A., 2010. A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. *Regul Toxicol and Pharmacol* 56, 67-81. https://doi.org/10.1016/j.yrtph.2009.09.006

Yang, C., Barlow, S. M., Muldoon Jacobs, K. L., Vitcheva, V., Boobis, A. R., Felter, S. P., Arvidson, K. B., Keller, D., Cronin, M., 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 chemical toxicol 109*(Pt 1), 170–193. https://doi.org/10.1016/j.fct.2017.08.043

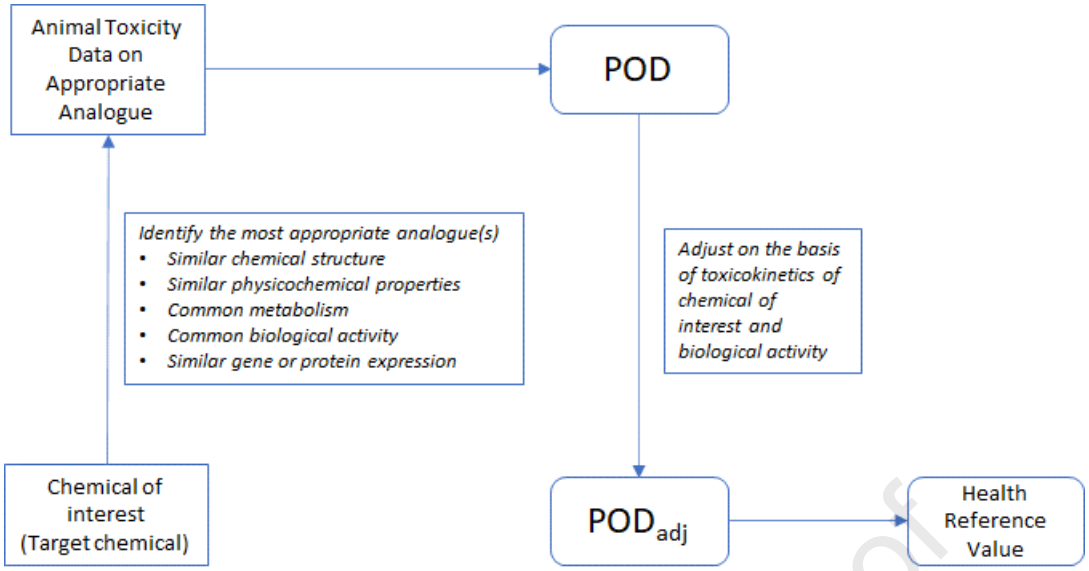
Figure Legends

Figure 1 (as appears in the NAS 2017 report) An approach to deriving health reference values when data on similar chemicals are available. Similarity can be based on such characteristics as chemical structure, physicochemical properties, metabolism, key events in biological pathways, or gene/protein expression; similarity of several characteristics increases confidence in the analogy. The point of departure (POD) of the appropriate analogue could be adjusted in this approach on the basis of toxicokinetic and potency differences between the chemical of interest and the analogue (e.g. with respect to biological activity such as receptor activation) or a relative potency factor could be used in the risk assessment. Relevant uncertainty factors would then be applied or models would be used to derive a health reference value. Accounting for uncertainty results in a determination of the level of confidence in the read-across, and would include consideration of the degree of similarity of the analogue to the chemical of interest and the extent and quality of both the *in vivo* data and the new approach methodology (NAM) data on the analogue and target chemical.

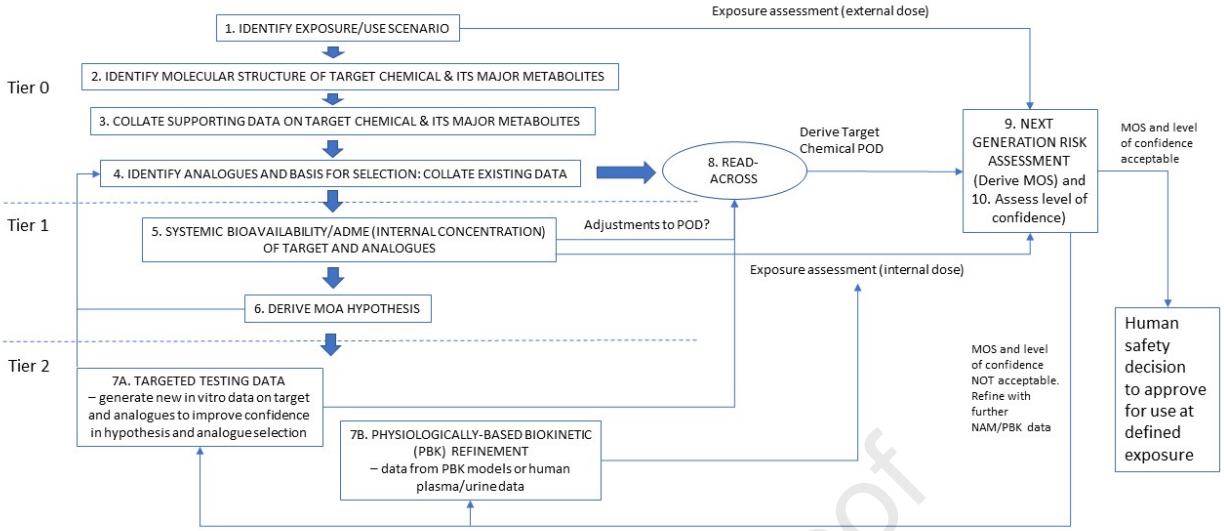
Figure 2 Nine principles of next generation risk assessment (NGRA) (Dent et al., 2018)

Figure 3 A tiered 10-step framework to enable a human safety decision to be made using NAMs and RAX, which in (a) diagrammatically builds on the SEURAT 1 workflow (Berggren et al., 2017) to perform a next generation risk assessment without new animal data; the steps are tabulated in (b).

Figure 4 The positioning of PBK when used in the context of refining exposure for a risk assessment (building on the concepts in Embry et al 2014).



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



b)

Tier 0
Step 1: Identify exposure/use scenarios for target chemical

Step 2: Identify molecular structure of target chemical

Step 3: Collate supporting data on target chemical and define data gap(s)

Step 4: Analogue(s) a) Identify, b) collate existing data, c) determine similarity hypothesis

End Tier 0 → Potential to move to Steps 8-10 if data are sufficient

Tier1
Step 5: Systemic bioavailability/ADME of target chemical and analogues

Step 6: Supporting a Similar Mode/Mechanism of Action (MoA) hypothesis

End Tier 1 → Potential to move to Steps 8-10 if data are sufficient

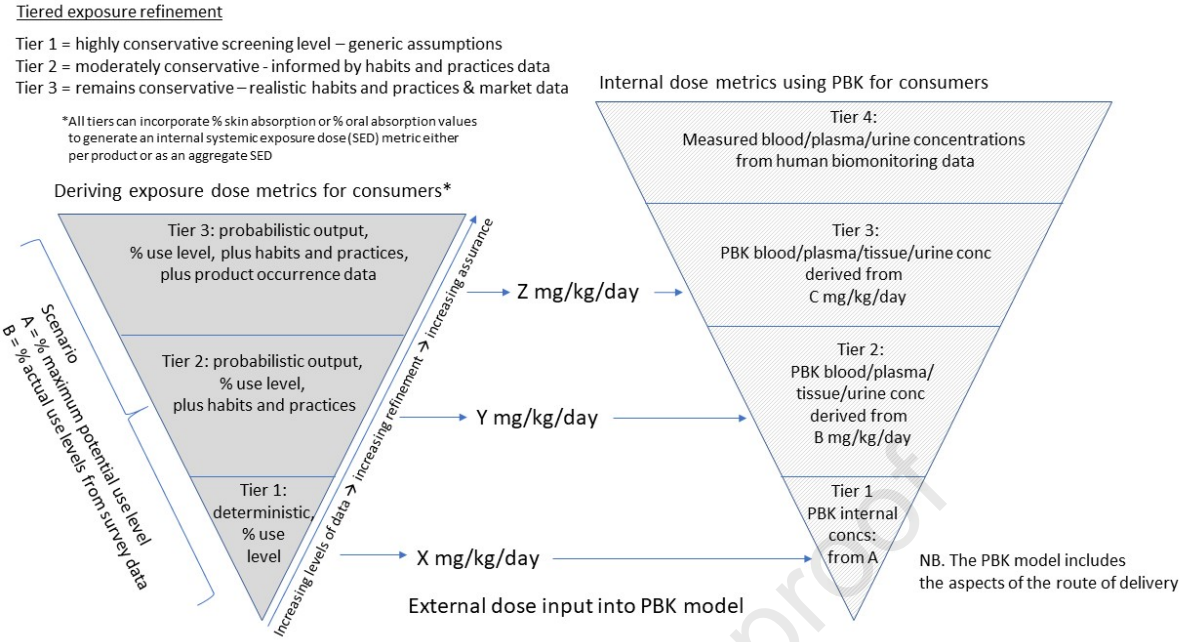
Tier 2
Step 7: a) Perform targeted testing to strengthen hypotheses and/or b) Biokinetic refinements of target chemical and analogues

The Assessment

Step 8: Performing a read-across (RAX) to derive a point of departure (POD)

Step 9: Performing a margin of safety (MOS) evaluation

Step 10 Assessing the level of confidence for establishing if the MOS is acceptable



AlexWhite et al – Highlights

- A 10-step framework for applying read-across (RAX) and novel approach methods (NAM) in cosmetics safety assessment
- Confidence in using RAX and NAM in cosmetics safety assessment by defining mode(s) of action in biological effect pathways
- Incorporating physiologically-based biokinetic (PBK) modelling to refine cosmetics ingredient exposure assessments
- Using NAMs for both toxicokinetics and toxicodynamics in tiered and integrated assessment

Funding Body – Alexander-White et al 'A 10-Step Framework for Use of Read-Across (RAX) in Next Generation Risk Assessment (NGRA) for cosmetics safety assessment'

This work was funded through the Long Range Science Strategy (LRSS) programme of Cosmetics Europe (https://www.lrsscosmeticseurope.eu).

Declaration of interests – Alex-White et al 'A 10-Step Framework for Use of Read-Across (RAX) in Next Generation Risk Assessment (NGRA) for cosmetics safety assessment'

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: