Lessons Learned from Read-Across Case Studies for Repeated-Dose Toxicity

Terry W. Schultz[1] and Mark T.D. Cronin[2*]

[1] The University of Tennessee, College of Veterinary Medicine, 2407 River Drive, Knoxville, TN 37996-4543 USA; [2] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, L3 3AF Liverpool, England

* Corresponding author.

E-mail addresses: tschultz@utk.edu (T.W. Schultz), m.t.cronin@ljmu.ac.uk (M.T.D. Cronin).

Disclaimer: The views expressed in this document are those of the authors and do not necessarily represent the official views of Cosmetics Europe or its Members.

Word Count for Abstract: 198 words

Word Count for Text: 4,727 words

Word Count for References: 1,555 words

ABSTRACT

A series of case studies designed to further acceptance of read-across predictions, especially for chronic health-related endpoints, have been evaluated with regard to the knowledge and insight they provide. A common aim of these case studies was to examine sources of uncertainty associated with read-across. While uncertainty is related to the quality and quantity of the read across endpoint data, uncertainty also includes a variety of other factors, the foremost of which is uncertainty associated with the justification of similarity and quantity and quality of data for the source chemical(s). This investigation has demonstrated that the assessment of uncertainty associated with a similarity justification includes consideration of the information supporting the scientific arguments and the data associated with the chemical, toxicokinetic and toxicodynamic similarity. 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. Data from New Approach Methodology(ies) including high throughput screening, *in vitro* and *in chemico* assay and *in silico* tools, may provide critical information needed to strengthen the toxicodynamic similarity rationale. In addition, it was shown that toxicokinetic (i.e., ADME) similarity, especially metabolism, is often the driver of the overall uncertainty.

Keywords: read-across; similarity; uncertainty; case studies; repeated-dose toxicity; regulatory acceptance.

Highlights:

- Read-across case studies for repeated-dose toxicity were evaluated
- Identification and definition of uncertainties in read-across is crucial
- The logic and data leading to a read-across prediction must be documented
- The similarity rationale of a read-across should be described transparently
- The roles of any endpoint specific and/or non-specific factors should be clarified

1. Introduction

Legislative requirements for the registration and safety assessment of chemicals, along with the need to obtain toxicological information without resorting to animal testing, have stimulated a more critical examination of read-across (RA). The concept of category formation, chemical grouping and RA is used to support chemical safety assessment by filling data gaps without the need for further *in vivo* testing (ECHA, 2014; OECD, 2014a; Stanton and Kruszewski, 2016). Historically, the fundamental assumptions of RA are that chemicals, which are similar in their structure, will have similar chemical properties and, thereby, have similar toxicokinetic and toxicodynamic properties (Cronin et al., 2013). A group of substances with similar toxicokinetic and toxicodynamic properties can be considered a toxicological meaningful category or a group of chemicals whose human health and/or environmental toxicological properties are likely to be similar or follow a regular pattern for a particular hazard. RA of toxic potencies based on such a category is a valuable approach to data gap filling, thus having a number of regulatory applications. Briefly, experimentally-derived toxicological properties from one or more source chemicals may be read across to fill the data gap for a target chemical, which is "similar" and for which an experimentally derived toxicological value is wanting and such prediction can be used for screening, priority setting, hazard assessment or risk assessment (Patlewicz and Fitzpatrick, 2016).

1.1. Background

Since the review of Cronin et al (2013), a number of papers have appeared that focus on modern-day RA. Many of these, including Blackburn and Stuard (2014), European Chemicals Agency (ECHA) (2015), Organisation for Economic Co-operation and Development (OECD) (2015) and Schultz et al. (2015), have put forward efforts to improve RA arguments and improve and innovate approaches (Batke et al., 2016; de Abrew et al.,

2016; Shah et al., 2016; van Ravenzwaay et al., 2016). More recently, Ball et al., (2016) summarised the state-of-the-art surrounding read-across, along with reasons relating to regulatory non-acceptance, and compiled relevant guidance under the heading of "Good Read-Across Practice"; Hartung (2016) described the concept of linking different types of data and tools under the umbrella of good read-across practices.

It is acknowledged RA is not a new concept (cf. Hanway and Evans, 2000), despite this, a number of challenges continue to impede its wider use. When applying RA to fill a toxicological data gap, a number of fundamental questions repeatedly arise (Schultz et al., 2014), including:

- Is it possible to form a robust group of chemicals (often referred to as a chemical category) which includes the target chemical?
- Is the category relevant to fill a data gap considering the toxicology of the endpoint under assessment?
- Are there appropriate toxicology studies of sufficiently high quality for the source chemical(s) to allow a meaningful RA?
- Are the uncertainties defined and are they acceptable in order to use the read across prediction(s) to fill the data gap(s) for a specific regulatory purpose?

To address these questions and others, a flexible strategy for developing and reporting a repeated-dose RA prediction was devised and applied in the case studies (Schultz et al., 2015). Briefly, this strategy focuses on the two main elements of a RA, namely:

- assessment of the similarity between source and target substance(s) and,
- assessment of the uncertainties in the RA process and ultimate prediction.

It is worth noting the publication of this strategy predates ECHA's Read-Across Assessment Framework (RAAF) (ECHA, 2015). Regardless of process, the standards for accepting a RA

prediction are likely to vary little, as the aim of a RA is to provide a prediction(s) that is (more or less) equivalent to that which would be obtained from the standard animal study.

In order to address at least some of the above questions, and to determine the suitability of RA to fill data gaps for repeated-dose toxicity (focussing on the oral route of exposure to the rat), Berggren et al. (2015) recommended that a series of case studies be conducted for the most likely RA scenarios. An additional recommendation was that each case study be evaluated in a two-step process. The initial step was to be a "traditional" RA using established *in vivo* data supplemented, as applicable, with conventional *in vitro* and classic structure-activity relationship information. The second iteration was to be a RA with the initial information and data supplemented with "New Approach Methodology" (NAM) data from high-throughput screening (HTS), novel *in vitro* methods and/or toxicogenomic assays.

Following an external review process, the findings of four case studies for the filling of data gaps for repeated-dose toxicity using RA have been published, covering a variety of RA scenarios. The RA case studies were all for 90 day rat repeated-dose toxicity and explored:

i) The suitability of 2-propen-1-ol as a read-across analogue for other short chain primary and secondary β-olefinic alcohols on the basis of similarity in metabolic transformation (Przybylak et al., 2017).

ii) The use of data for short-chain mono-alkylphenols to fill data gaps for other mono-alkylphenols on the basis of similarity in toxicokinetics and toxicodynamics (Mellor et al., 2017).

iii) An investigation of saturated 1-alkanols presumed to be of low toxicity and varying in toxicokinetics as a results of alkyl chain (assuming no branching on the alkyl chain) (Schultz et al., 2017a).

iv) Consideration of 2-alkyl-1-alkanols where branching of the alkyl chain may affect RA for low toxicity chemicals (Schultz et al., 2017b).

Whilst the reader is encouraged to examine the case studies (Przybylak et al., 2017; Mellor et al., 2017; Schultz et al., 2017a; Schultz et al., 2017b), a summary of the findings is presented in Table 1. As summarised in Table 1, the four RA case studies were evaluated in terms of the robustness of arguments and the uncertainty associated with the different elements of the category formation. It is important to note that these case studies were not performed for the purpose of regulatory submission, but to investigate the process of RA and how it could be improved. As such they provide a rich source of potential knowledge and learning for the development and direction of future RA studies. It is also acknowledged that various other RA case studies have been published (Blackburn et al., 2011; de Abrew et al., 2016; van Ravenzwaay et al., 2016) and, whilst they have not been evaluated explicitly in this investigation as they are based on different endpoints and approaches, there has been implicit learning from these.

131 Table 1. Summary of the main findings of the read-across case studies for repeated dose chronic toxicity

Chemical Category and Reference	Key Features of Similarity Amongst All Compounds in the Category in Terms of Structure, ADME and Toxicity	Summary of Weight(s) of Evidence To Support Read-Across for the Category	Conclusion Regarding Uncertainty
n-alkanols; C5-C13 (Schultz et al 2017a)	• Single OH group; C5-C13 chain length, straight-chain hydrocarbon scaffolding. • Absorbed from the gut; distributed in the blood in solution; first pass metabolism leads mainly to the corresponding carboxylic acid; subsequent mitochondrial β-oxidation to CO_2 • No systemic toxicity; no chemical reactivity or receptor-mediated interactions; nonpolar narcosis is a probable mode-of action.	Chemistry: High Toxicokinetics: Medium Toxicodynamics: *In vivo*: High *In vitro*: High Overall: High	Same as performing an OECD TG 408 test

2-alkyl-1-alkanols; C5-C13 (Schultz et al 2017b)	• Single OH, C5-C13 chain length with a 2-position C1-C3 hydrocarbon scaffolding. • Absorbed from the gut; distributed in the blood in solution; first past metabolism leads mainly to glucuronidation; subsequent elimination in the urine. • No systemic toxicity; no chemical reactivity or receptor-mediated interactions; probable mode-of action is nonpolar narcosis.	Chemistry: High Toxicokinetics: Medium Toxicodynamics: *In vivo*: High *In vitro*: High Overall: 2-ethyl- and 2-propyl-1 alkanols: High 2-methyl-1-alkanols: Medium	Same as performing an OECD TG 408 test
β-olefinic alcohols; C3 to C6 (Przybylak et al 2017)	• Single OH group; C3-C6 hydrocarbon scaffolding with β-vinylic moiety. • Absorbed from the gut; distributed in the blood in solution; first past metabolism leads to the corresponding α, β-unsaturated aldehyde or α, β-unsaturated ketone.	Chemistry: High Toxicokinetics: Medium Toxicodynamics: *In vivo*: Medium *In vitro*: High Overall:	Straight-chain β -olefinic alcohols: same as performing an OECD TG 408 test;

	• The corresponding α, β-unsaturated derivatives are the definitive electrophilic toxicants; likely mode-of action is Michael addition; *in vivo* potency is related to relative thiol reactivity.	Straight-chain β-olefinic alcohols: high Branched-chain β -olefinic alcohols: medium	branched-chain β -olefinic alcohols: 2-propen-1-ol is the worst case scenario
mono-substituted alkyl phenols; ≤ C4 substituent (Mellor et al 2017)	• Single (≤ C4) alkyl group on a phenolic ring scaffolding. • Absorbed from gut, distributed in the blood in solution, first past metabolism leads mainly to Phase II conjugation to glucuronides and sulphates; excreted in the urine. • No systemic toxicity; no chemical reactivity; no relevant receptor-mediated interactions; probable mode-of action is polar narcosis.	Chemistry: High Toxicokinetics: Medium Toxicodynamics: *In vivo*: High Overall: High	Same as performing an OECD TG 408 test

132

133

134 *1.2. The aim*

135 The present paper recapitulates with examples the lessons learned from the recent series of
136 case studies which illustrate specific issues associated with modern-day toxicological RA of
137 repeated dose toxicity. The case studies cited have the advantage of having undergone
138 external review prior to publication. We believe this summary of lessons has the potential of
139 furthering the acceptance of RA predictions, especially for predictions of NOAELs from
140 repeated-dose toxicity studies.

141

142 **2. Methods and Materials**

143 The findings reported here build on previous analyses, starting with guidance (ECETOC,
144 2012; ECHA, 2009, 2011; OECD, 2007, 2011, 2014a, 2015) on grouping of chemicals and
145 RA as well as other publications in this area (Ball et al., 2014, 2016; Cronin et al. (2013);
146 Blackburn and Stuard, 2014; Patlewicz et al., 2013a, 2013b, 2014, 2015; Patlewicz and
147 Fitzpatrick, 2016).

148 As stated in the introduction, the case studies from which the findings in this paper were
149 arrived at were Przybylak et al. (2017), Mellor et al. (2017) and Schultz et al. (2017a, 2017b).
150 Each case study is consistent with RA principles previously described (e.g., ECHA, 2013a,
151 2013b; OECD, 2015; Schultz et al., 2015). These four case studies, while developed by an
152 iterative effort of their authors, were extensively reviewed by various independent experts.

153

154 **3. Lessons Learned**

155 RA case studies have crucial evidentiary value in regulatory toxicology. Amongst other
156 things, they provide a means of illustrating how it may be possible to move from chemical-

157 by-chemical assessments based on animal testing to assessments by interpolation within a
158 toxicologically-relevant and mechanistically plausible chemical category. In the context of
159 this paper, case studies provided an opportunity to benchmark some of the lessons learned or
160 confirmed to use RA in a regulatory context.

161 *3.1. Today's read-across*

162 Early approaches to RA, e.g. identification of analogues with varying chain lengths (Hanway
163 and Evans, 2000; Patlewicz and Fitzpatrick, 2016) are increasingly being seen as simplistic
164 and trivial, especially for regulatory use. Today "chemical similarity" means more than
165 proving similarity in chemistry; it requires the category formation and RA process to be
166 transparent, reproducible and clearly documented. Specifically, key principles of biological,
167 as well as chemical, similarity need to be supported, where possible, by data and scientific
168 evidence. To improve the acceptability of any RA estimate, there is a need to justify the
169 prediction by explaining in a scientifically defensible manner how it was derived and why it
170 is justifiable for the intended purpose (e.g., priority setting, risk assessment).

171 Analysis of the four RA case studies confirmed that transparency should take the form of
172 reporting, in appropriate detail, all relevant data used both to establish similarity arguments
173 and the read across value. Such transparency can be accomplished in either a narrative
174 presentation or in tabular form – with advantages and disadvantages to both. The latter, while
175 leading to a shorter document, often requires the assessors to identify key information and/or
176 data from the tables. The former often leads to a longer document where, however, the same
177 information may be repeated.

178 In order to accept a RA prediction(s), the assessor needs to have confidence in the accuracy
179 of the prediction. This confidence can be attained in several ways. One method is to provide
180 sufficient information and data in the RA documentation such that the prediction can be

reproduced. To meet these requirements for transparency and reproducibility, clear documentation, especially by following a template or strategy (e.g., OECD 2015; Schultz et al., 2015, ECHA, 2015), is extremely helpful. Such methodologies provide "check lists" of the types of information critical to establishing confidence. As part of the "New Approach Methodologies" workshop one of the four cases studies on which this article is based was assessed, independently, within the context of the ECHA RAAF (ECHA, 2015), it was clear that the RAAF provided a structured means of determining the confidence in a RA. Further, this leads to the possibility of developing a template for developing RA arguments for regulatory use from the criteria detailed in the RAAF.

3.2. Uncertainties in RA

Sources of uncertainty in RA must be identified and addressed. Uncertainty includes a variety of elements which can be separated into:

- the uncertainty associated with the concept and completeness of the RA argument and,
- the uncertainty associated with the similarity justification (Schultz et al., 2015).

There are uncertainties associated with the basic tenet of the RA (i.e., the results of *in vivo* study/ies of the source substance(s) which are to be read across to the target analogue). The case studies confirmed that a key issue in any RA argument is to lower the overall uncertainty to an acceptable level; without high quality data for at least one source substance, it is not possible to conduct RA. With regard to risk assessment, the required level of uncertainty in the RA prediction should be similar to that associated with an appropriate *in vivo* test (e.g., OECD TG 408) and was assessed in the context of the template provided by Schultz et al. (2015). For other regulatory applications (e.g., screening and priority setting), higher levels of uncertainty may be acceptable. Thus, the first source of uncertainty that must be addressed is

related to the quality of the read across data (e.g., 90-day oral repeated-dose no observable adverse effect level (NOAEL)). Analysis of the data considered in the case studies indicated that such uncertainty may be best addressed using data from well-designed studies where the phenotypic expressions at the lowest observable adverse effect level (LOAEL) are reported e.g. for repeated-dose toxicity, a study which followed OECD TG 408. The quantity of *in vivo* data also affects uncertainty. It is intuitive that reading across from multiple source substances (i.e., a category approach), as in the 1-alkanol case study (Schultz et al., 2017a), has less uncertainty than reading across from one source substance (i.e., an analogue approach), as in the β-olefinic alcohol case study (Przybylak et al., 2017). Uncertainty within a RA is reduced by interpolation within a category bracketed by experimental *in vivo* data, as illustrated in the alkyl-phenol case study (Mellor et al., 2017), as opposed to extrapolation from data for one structural extreme of a category to another, as described in the 2-alkyl-1-alkanol case study (Schultz et al., 2017b).

Uncertainty associated with the similarity arguments is based on two interrelated rationales. Firstly, that the target and source substances are "sufficiently similar" to be toxicologically relevant. Secondly, "differences in similarity" are not relevant to the endpoint under consideration (Schultz et al., 2015). Confirming the latter was found to be a problematic task in the case studies. For example, in the 2-alkyl-1-alkanol case study (Schultz et al., 2017b), the authors believe the 2-methy-derivatives are "sufficiently similar" to the 2-ethyl- and 2-propyl-derivatives; however, the lack of *in vivo* repeated-dose toxicity data for a 2-methyl-derivative made it difficult to say with confidence that the structural difference does not preclude the inclusion of the methyl-substituted derivatives in a category defined with experimental data for 2-ethyl- and 2-propyl-substituted primary alcohols.

The assessment of uncertainty associated with a similarity justification is shown to include consideration of all information supporting the similarity argument. For example, for repeated-dose toxicity, the four case studies all revealed that not only are similarities in structure and chemical properties required, but also similarities associated with toxicokinetic and toxicodynamic properties. As such, data from *in vitro* assays, including new-methods (e.g., HTS) and *in silico* tools, often provide critical information needed to strengthen mechanistic and toxicodynamic similarity rationales. Establishing toxicokinetic similarity (i.e., ADME properties), especially for metabolism and metabolic rates, is often the driver of the overall uncertainty (Ball et al., 2014; Hand et al., 2017) and was shown to be a key factor in the separation of the saturated aliphatic primary alcohols into two categories for RA (see Schultz et al., 2017a, 2017b).

3.3. Toxicological meaningful categories for repeated-dose toxicity

Forming a toxicological meaningful category for longer-term hazards such as repeated-dose toxicity becomes a daunting task (Berggren et al., 2015). Data reported for repeated-dose hazards are elaborate, typically based on a defined vocabulary list of possible symptoms. For example, ToxRefDB (US EPA, 2008) includes *in vivo* data for 309 structures and lists more than 22,000 vocabulary items; whilst some are quantitative, most are qualitative (e.g., positive versus negative; increase versus decrease). Symptoms reported include clinical chemistry, haematological and urinalysis, as well as gross and microscopic pathological findings. It is intuitive that these symptoms cannot all be related to the same *in vivo* effect. Since exposure for chronic effects is over a longer duration (e.g., 28 days or longer in the case of repeat-dose toxicity), the *in vivo* damage is likely to be cumulative. Thus, the reported values such as a NOAEL or LOAEL vary as to incidence, target organ, severity etc. (Martin et al., 2009). Whilst limitations to forming a toxicologically meaningful category for quantitative RA include the availability of suitable *in vivo* data to be read across and the lack

of toxicologically-relevant *in vitro* or NAM data to support mechanistic plausibility, the major limitation to using RA for repeated dose endpoints is often the lack of toxicokinetics understanding and data.

3.4. Mechanistic plausibility

While not always possible, stating and documenting mechanistic plausibility improves the likelihood of a RA prediction being accepted. This is especially true if mechanistic plausibility can be linked to a mode of toxic action or an adverse outcome pathway (AOP) (Ellison et al., 2016). An adverse outcome pathway is a description of plausible causal linkages, illustrating how a molecular initiating event may lead to the key biochemical, cellular, physiological behavioural etc. responses resulting in an apical effect; it thus characterises the biological cascade across the different levels of biological organisation (OECD, 2013; OECD 2014b).

As seen with the case studies for n-alkanols and 2-alkyl-1-alkanols (Schultz et al., 2017a; 2017b), even incomplete mechanistic understanding in the form of presumptive AOPs has value in establishing toxicological meaningful categories. Moreover, presumptive AOPs provide a means of linking *ex vivo, in vitro* and *in chemico* effects to the apical *in vivo* endpoint of interest. As demonstrated in the β-olefinic case study (Przybylak et al., 2017), confidence in mechanistic plausibility can be increased by the use of toxicologically-relevant alternative methods data. In the latter case, the *in vivo* data for a single source substance are supported by *ex vivo* data for five category members, including the source substance, as well as *in chemico* data for 16 category members, again including the source substance. In this way, NAM data have contributed to mechanistic understanding and hence supported the hypothesis of category membership.

3.5. Endpoint specificity

277 Predictions from RA are more likely to be acceptable when undertaken on an endpoint-by-
278 endpoint basis. The case studies for 1-alkanols and 2-alkyl-1-alkanols (Schultz et al., 2017a,
279 2017b) demonstrated that for acute oral rat toxicity, measured as the LD50 (mg/kg), the two
280 alkanols sub-classes form a single category. Specifically, there is a similar mode of toxic
281 action, similar Toxicity Forecaster database (ToxCast) molecular fingerprints and similar
282 experimental LD50 values of ≈3000 mg/kg bw. Thus, both sub-classes belonged to the same
283 category for rat acute oral toxicity and experimental results (i.e., the LD50 value) can be read
284 across to untested analogues with acceptable uncertainty. In contrast, for read-across of 90-
285 day oral repeated-dose toxicity endpoint, expressed as the NOAEL values (mg/kg bw/d), the
286 n-alkanols and the 2-alkyl-1-alkanols formed two separate categories. Specifically, whilst
287 similar in mode of toxic action with similar ToxCast results, the NOAEL values (≈1000 and
288 ≈125 mg/kg bw/d for n-alkanols and the 2-alkyl-1-alkanols, respectively) are dissimilar. In
289 this case read across for rat oral repeated-dose toxicity can be achieved with acceptable
290 uncertainty only after appropriate sub-categorisation into the different chemical sub-classes
291 (see Schultz et al., 2017a, 2017b).

292 *3.6. Rationale for grouping substances*

293 RA approaches have been developed on an over-arching rationale for grouping substances
294 based on molecular structure and chemical properties (Cronin et al., 2013). The case studies
295 have, however, demonstrated that these similarities in chemistry alone are often not sufficient
296 to justify a RA prediction. This is especially the case for sub-chronic and chronic health
297 effects, where multiple dosing may lead to different toxicokinetic and toxicodynamic
298 properties (Schultz et al., 2015). Further information that is often required typically includes
299 that related to toxicokinetic and toxicodynamic properties e.g., metabolism, clearance,
300 mechanistic plausibility etc. The case studies for the 1-alkanols and 2-alkyl-1-alkanols
301 (Schultz et al., 2017a, 2017b) demonstrated that the over-arching rationale for grouping is the

same, i.e. highly similar chemistry and similar mechanistic plausibility in the form of an anaesthetic-like mode of toxic action. Despite this, key toxicokinetics parameters are different. Specifically, whilst absorption and distribution are highly similar for both groups of saturated alcohols, metabolism and elimination are different. 1-Alkanols, such as 1-octanol, are excreted mainly (>90%) as CO_2, and to a lesser extent as n-glucuronide in the urine (Schultz et al., 2017a). In contrast, experimental data reveal that the major pathways of metabolism of branched saturated alcohols, such as 2-alkyl-1-alkanols, lead to conjugation with glucuronic acid. In addition, there is often oxidation of the alcohol group, as well as side-chain oxidation (Schultz et al., 2017b). Thus, whilst there is structural similarity within the alkanols, sub-categorisation is required to facilitate efficient RA.

3.7. Weights-of-Evidence (WoE)

The consideration of all relevant information used to undertake and support a RA can be achieved through Weight(s)-of-Evidence (WoE). Clearly, increased WoE has the possibility to reduce uncertainties both in relation to similarity justifications and the completeness of the read-across argument. The increased WoE can take the form of using different types of *in vivo* information and data (Schultz et al., 2017a, 2017b) or using *in vitro* and/or *in chemico* information and data (Mellor et al., 2017; Przybylak et al., 2017) – as such it can be seen as an extension of the support that can be provided for mechanistic plausibility.

The case study for 1-alkanols (Schultz et al., 2017a) demonstrated that uncertainty associated with RA predictions was reduced by increasing WoE through the addition of information from *in vivo* data. Specifically, the uncertainty associated with the RA of a 90-day rat oral repeated-dose NOAEL may be reduced following the inclusion of *in vivo* information from derivatives tested with a related protocol (e.g., OECD TG 408 vs OECD TG 422 studies) where qualitative and quantitative similarity are established. In addition, the β-olefinic alcohols case study (Przybylak et al., 2017), demonstrated that non-animal test data increase

the WoE for the RA justification through the results from toxicologically-relevant alternative methods. This increase in WoE can be in the form of relevant data from both the source and target chemicals, as well as relevant data for other substances within the applicability domain.

The overall WoE may also be improved by having information for chemicals that may, in terms of mechanistic plausibility, be considered to be outside the category. The β-olefinic alcohol case study (Przybylak et al., 2017), demonstrated that saturated alcohols, which were outside the domain of the RA, exhibited test results for *ex vivo* and *in chemico* endpoints markedly different from results for β-olefinic alcohols. Thus, the most likely mechanism of toxic action (i.e., alcohol dehydrogenase mediated metabolism leading to a Michael-acceptor electrophile) is limited to β-olefinic alcohols. These consistent differences also add to the WoE.

3.8. Using New Approach Methodology data

Relevant and reliable NAM data are useful in reducing the uncertainty in toxicodynamics and improve mechanistic understanding (ECHA, 2016). The use of data from NAMs will assist in the acceptance of a "low/no toxic" RA prediction where a higher level of certainty is likely to be required (Schultz et al., 2017a).

In silico methods are the easiest, most rapid and often the cheapest NAM and are appealing as a good alternative first approach. However, they are useful only when there is confidence that *in silico* models are of high quality and have been applied correctly. In all four case studies, *in silico* models were used to establish similarities in physico-chemical properties, as profilers for possible toxicophores, and to examine potential metabolic similarity.

Several of the case studies demonstrated that HTS (i.e., ToxCast results) can assist in establishing the most likely pathway leading to an *in vivo* adverse outcome. Typically, these

data were used to support a mode of toxic action developed from non-mammalian data or a presumptive AOP (see Schultz et al., 2017a; Mellor et al., 2017).

In vitro assays which express a particular pathway and associated pathologies were found to be useful to link relevant *in vitro* data to the apical endpoint predicted in the RA. In this way, mechanistically fit-for-purpose data can reduce uncertainty and increase the WoE. For example, in the β-olefinic alcohol case study (Przybylak et al., 2017), metabolism is presumed to lead to derivatives that cause oxidative stress thus leading to narcosis and apoptosis that, in principle, lead to *in vivo* fibrosis. The incorporation of NAMs data into RA arguments allowed for the addition of relatively rapid and inexpensive hypothesis-driven testing and evaluation. This has the advantage of performing targeted, rather than universal, tests.

In the near future, new *in vitro* systems based on multiple cell-to-cell interactions and fit-for-purpose based mechanistic reasoning are likely to add confidence to RA predictions. For example, 3D co-culture systems, such as those reported by Wink et al. (2014) and Ramaiahgari et al. (2014) consisting of a HepG2 BAC-GFP reporter system, or that of Leite and co-workers (2015) consisting of hepatic organoids (of human hepatocyte-like cells (HepaRG) and primary human hepatic stellate cells) were used to measure stress response activation and fibrosis as proposed in the β-olefinic alcohol case study (Przybylak et al., 2017).

3.9. Size of category

The development and evaluation of the case studies also highlighted the balance required between reducing uncertainty by restricting the size of a category i.e. its applicability domain and making it a meaningful size. At the extremes category definition may be very broad e.g. aliphatic alcohols, or highly specific restricting it to a very small number of members. The

case studies demonstrated that, in reality, neither is ideal for RA. Broad definitions of applicability domains often do sharply increase the uncertainty associated with RA predictions for some substances within the domain, whilst very narrow definitions make decrease the practical utility and make it more difficult to build up a body of evidence. For example, in the initial RA evaluation of the β-unsaturated alcohols (Przybylak et al., 2017), a broader category including the β-acetylenic alcohols (e.g., 1-propyn-3-ol) was considered. Subsequently, due to toxicokinetic uncertainty, these derivatives were eliminated from consideration which had the effect in decreasing uncertainty, but reduced the breadth of the applicability domain of the category. Furthermore, in the β-olefinic alcohol case study, the single source substance, allyl alcohol, is unique and is effectively a category on its own. While all α,β-unsaturated carbonyl compounds derived from hepatic metabolism of primary and secondary β-olefinic alcohol readily react with glutathione (Przybylak et al., 2017), there is less uncertainty associated with straight-chain alcohols (e.g., 1-alken-3-ols and 2-alken-1-ols) than with branched alcohols (2-methyl-2-alken-1-ols, 3-methyl-2-alken-1-ols). Accordingly, as noted by Przybylak et al. (2017), uncertainty may be better addressed via sub-categorisation. In the 2-alkyl-1-alkanol study (Schultz et al., 2017b), there is greater uncertainty associated with the 2-methyl-substituted derivatives. Whilst they are considered within the domain of the RA, there are no *in vivo* experimental data supporting their inclusion; thus, greater uncertainty.

4. Discussion

A basic understanding of what a RA to fill data gaps for hazard and / or risk assessment should look like, to garner regulatory acceptance, is rapidly taking shape (Teubner and Landsiedel, 2015; Ball et al., 2016). As noted by OECD, the lessons learned from the

cooperative review of case studies on grouping methods (such as RA) have increased experience in the application of these approaches (OECD, 2016). However, it has also been recognised that more case studies are needed for developing general guidance (OECD, 2016). The evaluation of the series of case studies on which this article is based has indicated that the acceptability of a RA prediction is often dependent on the evaluator's sense of confidence in the documentation and evidence provided. High confidence is associated with RAs when there is strong proof that the prediction is valid i.e., low uncertainty. The case studies have shown that improvement in the acceptance of RA predictions is accomplished in several ways. Firstly, high quality *in vivo* endpoint data are essential to anchor any RA; higher confidence is linked to qualitative and quantitative consistency with more than a single source substance. Secondly, it is essential to establish the adequacy and reliability associated with the underlying hypothesis of chemical and / or biological similarity. Higher confidence is linked to arguments built on toxicokinetics and toxicodynamics, support by an assessment of similarities in chemistry rather than being driven by chemical similarity alone. Thirdly, the depth and breadth of supporting information is important; higher confidence is linked to studies with increased WoE which is typically provided by information form non-animal methods. Fourthly, higher confidence is associated with strong evidence of mechanistic plausibility. Finally, higher confidence is associated with supporting evidence provided by hypothesis-based testing, especially with fit-for-purpose *in vitro* and *in chemico* assays. The latter come with the caveat that these sources of information meet reliability issues (e.g., repeatability and reproducibility) often need for regulatory acceptance.

4.1. Summary

RA arguments are best established:

1. in a context-dependent manner (one size does not fit all),

2. on one-to-one (analogue) or many-to-one (category) basis rather than a one to many or a many to many arguments, and
3. on the basis of a single well-defined endpoint (e.g., time, species, route of exposure, etc.).

In conclusion, addressing uncertainty in a RA prediction is central to regulatory acceptance. For some endpoints (e.g., ecotoxicity, skin sensitisation, genotoxicity), useful frameworks are available, whilst for other endpoints, especially those related to chronic health, further work is needed.

5. Acknowledgements

TWS was funded by a consultancy agreement with Cosmetics Europe, the Personal Care Association. MTDC acknowledges funding from the COSMOS Project which was funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 266835 and Cosmetics Europe.

6. References

Ball, N., Bartels, M., Budinsky, R., Klapacz, J., Hays, S., Kirman, C. and 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. Pharmacol. 68: 212-221.

Ball, N., Cronin, M.T., Shen, J., Adenuga, M.D., Blackburn, K., Booth, E.D., Bouhifd, M., Donley, E., Egnash, L., Freeman, J.J., Hastings, C., Juberg, D.R., Kleensang, A., Kleinstreuer, N., Kroese, E.D., Luechtefeld, T., Maertens, A., Marty, S., Naciff, J.M.,

445 Palmer, J., Pamies, D., Penman, M., Richarz, A.-N., Russo, D.P., Stuard, S.B.,
446 Patlewicz, G, van Ravenzwaay. B., Wu, S., Zhu, H. and Hartung, T. 2016. Toward
447 Good Read-Across Practice (GRAP) guidance. ALTEX 33: 149-166.

448 Batke, M., Gutlein, M., Partosch, F., Gundert-Remy, U., Helma, C., Kramer, S., Maunz, A.,
449 Seeland, M. and Bitsch, A. 2016. Innovative strategies to develop chemical categories
450 using a combination of structural and toxicological properties. Front. Pharmacol. 7:
451 321.

452 Berggren, E., Amcoff, P., Benigni, R., Blackburn, K. Carney, E. Cronin, M., Deluyker, H.,
453 Gautier, F., Judson, R.S., Kass, G.E.N., Keller, D., Knight, D., Lilienblum, W.,
454 Mahony, C., Rusyn, I., Schultz, T., Schwarz, M., Schüürman, G., White, A., Burton,
455 J., Lostia, A., Munn, S., and Worth, A. 2015. Chemical safety assessment using read-
456 across: How can novel testing methods strengthen evidence base for decision-
457 making? Environ. Health Perspect. 123: 1232-1240.

458 Blackburn, K., Bjerke, D. Daston, G., Felter, S. Mahony, C., Naciff, J., Robison, S. and Wu,
459 S., 2011. Case studies to test: A framework for using structural, reactivity, metabolic
460 and physicochemical similarity to evaluate the suitability of analogs for SAR-based
461 toxicological assessments. Regul. Toxicol. Pharmacol. 60: 120-135.

462 Blackburn, K. and Stuard, S.B., 2014. A framework to facilitate consistent characterization of
463 read across uncertainty. Reg. Toxicol. Pharmacol. 68, 353-362.

464 Cronin, M.T.D., Madden, J.C., Enoch S.J. and Roberts, D.W., (Eds.) 2013. Chemical
465 Toxicity Prediction: Category Formation and Read-Across Applications. The Royal
466 Society of Chemistry, Cambridge.

467 de Abrew, K.N., Kainkaryam, R.M., Shan, Y.Q.K., Overmann, G.J., Settivari, R.S., Wang,
468 X.H., Xu, J., Adams, R.L., Tiesman, J.P., Carney, E.W., Naciff, J.M. and Daston,

G.P. 2016. Grouping 34 chemicals based on mode of action using connectivity mapping. Toxicol. Sci. 151: 447-461.

Ellison, C.M., Piechota, P., Madden, J.C., Enoch, S.J. and Cronin, M.T.D. 2016. Adverse Outcome Pathway (AOP) informed modeling of aquatic toxicology: QSARs, read-across, and interspecies verification of modes of action. Environ. Sci. Technol. 50: 3995-4007.

European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), 2012. Category Approaches, Read-across, (Q)SAR. Technical Report No 116. ECETOC, Brussels, Belgium.

European Chemicals Agency (ECHA), 2009. Practical Guide 6: How to Report Read-Across and Categories, ECHA, Helsinki, ECHA-10-B-11-EN.

European Chemicals Agency (ECHA), 2011. The Use of Alternatives to Testing on Animals for the REACH Regulation 2011, ECHA, Helsinki, ECHA-11-R-004.2-EN.

European Chemicals Agency (ECHA), 2013a. Grouping of Substances and Read-Across Approach. Part 1: Introductory Note, ECHA, Helsinki, ECHA-13-R-02-EN.

European Chemicals Agency (ECHA), 2013b. Read-Across Illustrative Example. Part 2. Example 1 – Analogue Approach: Similarity Based on Breakdown Products, ECHA, Helsinki, ECHA-13-R-03-EN.

European Chemicals Agency (ECHA), 2014. The Use of Alternatives to Testing on Animals for the REACH Regulation - Second report under Article 117(3) of the REACH Regulation. http://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2014_en.pdf .

492 European Chemicals Agency (ECHA), 2015. Read-Across Assessment Framework (RAAF).
493 European Chemicals Agency; Helsinki, Finland.
494 http://echa.europa.eu/documents/10162/13628/raaf_en.pdf.

495 European Chemicals Agency (ECHA), 2016. New Approach Methodologies in Regulatory
496 Science: Proceedings of a scientific workshop. Helsinki, Finland.
497 https://echa.europa.eu/documents/10162/21838212/scientific_ws_proceedings_en.pdf
498 /a2087434-0407-4705-9057-95d9c2c2cc57

499 Hand, L.H., Richardson, K., Hadfield, S.T., Whalley, S., Rawlinson, P. and Booth, E.D.
500 2017. Use of read-across to simplify the toxicological assessment of a complex
501 mixture of lysimeter leachate metabolites on the basis of chemical similarity and
502 ADME behavior. Regul. Toxicol. Pharmacol. 83: 109-116.

503 Hanway, R.H. and Evans, P.F. 2000. Read-across of toxicological data in the notification of
504 new chemicals. Toxicol. Lett. 116 (Suppl. 1): 61.

505 Hartung, T. 2016. Making big sense from big data in toxicology by read-across. ALTEX 33:
506 83-93.

507 Leite, S.B., Roosens, T., El Taghdouini, A., Mannaerts, I., Smout, A.J., Najimi, M., Sokal, E.,
508 Noor, F., Chesne, C. and van Grunsven, L.A. 2015. Novel human hepatic organoid
509 model enables testing of drug-induced liver fibrosis in vitro. Biomaterials 78: 1-10.

510 Martin, M.T., Judson, R.S., Reif, D.M., Kavlock, R.J. and Dix, D.J. 2009. Profiling chemicals
511 based on chronic toxicity results from the U.S. EPA ToxRef Database. Environ.
512 Health Perspect. 117: 392-399.

513 Mellor, C.L., Schultz, T.W., Przybylak, K.R., Richarz, A.-N, and Cronin, M.T.D. 2017.
514 Read-across for rat oral gavage repeated-dose toxicity for short-chain mono-
515 alkylphenols: A case study. Comput. Toxicol. 2: 1-11.

Organisation for Economic Co-operation and Development (OECD), 2007. Guidance on Grouping of Chemical. Environment Health and Safety Publications, Series on Testing and Assessment No. 80, Report No. ENV/JM/MONO(2007)28.

Organisation for Economic Co-operation and Development (OECD) 2009. Report of the Expert Consultation to Evaluate an Estrogen Receptor Binding Affinity Model for Hazard Identification, Series on Testing and Assessment, No. 111. ENV/ JM/MONO(2009)33.

Organisation for Economic Co-operation and Development (OECD), 2011. Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories. Environment Health and Safety Publications, Series on Testing and Assessment No. 138. ENV/JM/MONO(2011)8.

Organisation for Economic Co-operation and Development (OECD) 2013. Guidance Document for Developing and Assessing Adverse Outcome Pathways (AOPs). Environment Health and Safety Publications, Series on Testing and Assessment No. 184. ENV/JM/MONO(2013)6,.

Organisation for Economic Co-operation and Development (OECD) 2014a. Guidance on Grouping of Chemicals: Second Edition. Environment Health and Safety Publications, Series on Testing and Assessment No.194. ENV/JM/MONO(2014)4.

Organisation for Economic Co-operation and Development (OECD) 2014b. Users' Handbook Supplement to the Guidance Document for Developing and Assessing AOPs. ENV/JM/MONO(2013)6.

Organisation for Economic Co-operation and Development (OECD) 2014c. Draft Outline of Future Cooperative Work on the Hazard Assessment of Chemicals. 52nd Joint

539 Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides
540 and Biotechnology. ENV/JM/MONO(2014)100.

541 Organisation for Economic Co-operation and Development (OECD) 2015. Guidance
542 Document on the Reporting of Integrated Approaches to Testing and Assessment
543 (IATA). ENV/JM/HA(2015)7.

544 Organisation for Economic Co-operation and Development (OECD) 2016. Report on
545 Considerations from Case Studies on Integrated Approaches for Testing and
546 Assessment (IATA), Series on Testing & Assessment No. 250.
547 ENV/JM/MONO(2016)48

548 Patlewicz, G. and Fitzpatrick, J.M. 2016. Current and future perspectives on the
549 development, evaluation, and application of in silico approaches for predicting
550 toxicity. Chem. Res. Toxicol. 29: 438-451.

551 Patlewicz, G.Y., Ball, N., Booth, E.D., Hulzebos, E., Zvinavshe, E. and Hennes. C. 2013a.
552 Use of category approaches, read-across and (Q)SAR: general considerations. Regul.
553 Toxicol. Pharmacol. 67: 1-12.

554 Patlewicz, G.Y., Roberts, D.W., Aptula, A., Blackburn, K. and Hubesh, B. 2013b. Workshop:
555 Use of "read-across" for chemical safety assessment under REACH. Regul. Toxicol.
556 Pharmacol. 65: 226-228.

557 Patlewicz, G., Ball, N., Becker, R.A., Booth, E.D., Cronin, M.T.D., Kroese, D., Steup, D.,
558 van Ravenzwaay, B. and Hartung, T. 2014. Read-across approaches –
559 misconceptions, promises and challenges ahead. ALTEX – Altern. An. Exp. 31: 387-
560 396.

561 Patlewicz, G., Ball, N., Boogaard, P.J., Becker, R.A. and Hubesch, B. 2015. Building
562 scientific confidence in the development and evaluation of read-across. Regul.
563 Toxicol. Pharmacol. 72: 117-133.

564 Przybylak, K.R., Schultz, T.W., Richarz, A.-N., Mellor, C.L., Escher, S.E. and Cronin,
565 M.T.D. 2017. Read-across of 90-day rat oral repeated-dose toxicity: A case study for
566 selected β-olefinic alcohols. Comput. Toxicol. 1: 22-32.

567 Ramaiahgari, S.C., den Braver, M.W., Herpers, B., Terpstra, V., Commandeur, J.N., van de
568 Water, B. and Price, L.S. 2014. A 3D in vitro model of differentiated HepG2 cell
569 spheroids with improved liver-like properties for repeated dose high-throughput
570 toxicity studies. Arch. Toxicol. 88: 1083-1095.

571 Schultz, T.W., 2014. Chapter 2.6. Read-across as a basis for one of the SEURAT-1 proof-of-
572 concepts and an overview of the outcome of the SEURAT-1 read-across workshop, in
573 Gocht, T. and Schwarz, M. (Eds.), Toward the Replacement of in vivo Repeated Dose
574 Systematic Toxicity Testing. Volume 4: Implementation of the Research Strategy.
575 COACH Consortium, Paris, pp. 72-80.

576 Schultz, T.W., Amcoff, P., Berggren, E., Gautier, F., Klaric, M., Knight, D. J., Mahony, C.,
577 Schwarz, M., White, A. and Cronin, M.T.D. 2015. A strategy for structuring and
578 reporting a read-across prediction of toxicity. Reg. Toxicol. Pharmacol. 72: 586-601.

579 Schultz, T.W., Przybylak, K.R., Richarz, A.-N., Mellor, C.L., Escher, S.E., Bradbury, S.P.
580 and Cronin, M.T.D. 2017a. Read-across for 90-day rat oral repeated-dose toxicity for
581 selected n-alkanols: A case study. Comput. Toxicol. 2: 12-19.

582 Schultz, T.W., Przybylak, K.R., Richarz, A.-N., Bradbury, S.P. and Cronin, M.T.D. 2017b.
583 Read-across for 90-day rat oral repeated-dose toxicity for selected 2-alkyl-1-alkanols:
584 A case study. Comput. Toxicol. 2: 28-38.

585 Shah, I., Liu, J., Judson, R.S., Thomas, R.S. and Patlewicz, G. 2016. Systematically
586 evaluating read-across prediction and performance using a local validity approach
587 characterized by chemical structure and bioactivity information. Regul. Toxicol.
588 Pharmacol. 79: 12-24.

589 Stanton, K. and Kruszewski, F.H. 2016. Quantifying the benefits of using read-across and in
590 silico techniques to fulfil hazard data requirements for chemical categories. Regul.
591 Toxicol. Pharmacol. 81: 250-259.

592 Teubner, W. and Landsiedel, R. 2015. Read-across for hazard assessment: The ugly duckling
593 is growing up. ATLA 43: 67-71.

594 U.S. Environmental Protection Agency (US EPA), 2008. ToxRefDB Home.
595 http://www.epa.gov/ncct/toxrefdb/

596 van Ravenzwaay, B., Sperber, S., Lemke, O., Fabian, E., Faulhammer, F., Kamp, H., Mellert,
597 W., Strauss, V., Strigun, A., Peter, E., Spitzer, M. and Walk, T. 2016. Metabolomics
598 as read-across tool: A case study with phenoxy herbicides. Regul. Toxicol.
599 Pharmacol. 81: 288-304.

600 Wink, S., Hiemstra, S., Huppelschoten, S., Danen, E., Niemeijer, M., Hendriks, G., Vrieling,
601 H., Herpers, B. and van de Water, B. 2014. Quantitative high content imaging of
602 cellular adaptive stress response pathways in toxicity for chemical safety assessment.
603 Chem. Res. Toxicol. 27: 338-355.

604

605