Navigating through the minefield of read-across frameworks: A commentary perspective

Grace Patlewicz[a*], Mark T.D. Cronin[b], George Helman[a,c], Jason C. Lambert[d], Lucina E. Lizarraga[d], Imran Shah[a]

[a]National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency (US EPA), 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA

[b]School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

[c]Oak Ridge Institute for Science and Education (ORISE), 1299 Bethel Valley Road, Oak Ridge, TN 37830, USA.

[d]National Center for Evaluation Assessment (NCEA), US Environmental Protection Agency (US EPA), 26 West Martin Luther King Dr, Cincinnati, OH 45268, USA

*Correspondence: Grace Patlewicz

Tel: +1 919 541 1540

Email: patlewicz.grace@epa.gov

Running title: Read-across frameworks – An end-user perspective

Abstract

Read-across is a popular data gap filling technique used within analogue and category approaches for both regulatory and product stewardship purposes. In recent years there have been many efforts focused on the challenges involved in read-across development, its scientific justification and documentation for both chemical hazard and risk assessment purposes. Here, we summarise a selection of the read-across frameworks published in technical guidance documents or in the literature, and review their respective similarities and differences. There was a great deal of consensus between the different frameworks in terms of the general steps outlined and the similarity contexts considered although the terminology, decision context (chemical hazard and/or risk assessment purposes) and scope varied. A harmonised hybrid framework is proposed to help reconcile the common guiding principles and steps of the read-across process which should be helpful in expanding the scope and decision context of the existing frameworks. This harmonised framework is also intended to illustrate where generalised and systematic read-across approaches taking into consideration new approach methodology (NAM) information can be applied.

Keywords

chemical hazard and risk assessment; read-across development framework; read-across assessment framework; harmonised hybrid workflow; new approach methodology (NAM); generalised read-across (GenRA)

Highlights

- Read-across development and assessment frameworks are reviewed
- Similarities and differences of the frameworks are highlighted
- A harmonised hybrid framework is proposed
- Harmonised hybrid framework highlights where New Approach Methodologies (NAM) fit

Abbreviations

(ADME) Absorption, Distribution, Metabolism, Excretion; (AOPs) Adverse Outcome Pathways; (AUC) Area under the Curve; (AEs) Assessment Elements; (AOs) Assessment Options; (BMDL) Benchmark Dose Level; (CAAT) Center for Alternative to Animal Testing; (Cefic-LRI) European Chemistry Industry Council's Long Range Initiative; (CLP) Classification Labelling and Packaging regulation; (ECHA) European Chemicals Agency; (ECETOC) European Centre for Ecotoxicology and Toxicology of Chemicals; (EC JRC) European Commission's Joint Research Centre; (EU) European Union; (EU TCNES) EU Technical Committee for Existing Chemicals; (EU TC C&L) EU Technical Committee for Classification and Labelling; (HC) High Content; High Production Volume (HPV); (HT) High Throughput; (IATA) Integrated Approaches to Testing and Assessment; (log D) the distribution coefficient; (log Kow) the log of the octanol-water partition coefficient; (LOAEL)Lowest Observed Adverse Effect Level; (MIE) Molecular Initiating Event; (MoA) Mode of Action; (MW) molecular weight; (NAM) New Approach Methodologies; (NOAEL) No Observed Adverse Effect Level; (OECD) Organisation for Economic Co-operation and Development; (pKa) the acid dissociation coefficient; (POD) Point of Departure; (PPRTVs) Provisional Peer Reviewed Toxicity Values; (QSARs) Quantitative Structure-Activity Relationships; (RAAF) Read-Across Assessment Framework; (REACH) Registration Evaluation Authorisation and restriction of Chemicals; (ROC) Receiver Operating Characteristic; (RPF) Relative Potency Factor; (RMSE)Root Mean Square Error; (SAR) Structure-Activity Relationship; SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing); (TEF) Toxicity Equivalence Factor; (US EPA) United States Environmental Protection Agency; (UVCBs) Substances of unknown or variable composition, complex

reaction products or biological materials; (WOE) Weight of Evidence; (WPHA) Working Party on Hazard Assessment

1. Introduction

1.1 Background context

The legislative landscape particularly in Europe (e.g. Registration Evaluation Authorisation and restriction of CHemicals (REACH) [1], Classification Labelling and Packaging regulation (CLP) [2], Cosmetics Directive [3]) has stipulated that information requirements for hazard and risk assessment should be addressed without recourse to animal testing. As there are thousands of data-poor or toxicologically uncharacterised chemicals in commerce, this has invariably shined a light on read-across as a convenient and efficient data gap filling technique. Yet, this has also prompted a more critical examination of its application and scientific justification. Whilst read-across is traditionally anchored with conventional *in vivo* and *in vitro* data, concerted efforts are starting to be made to exploit high throughput (HT) and high content (HC) screening data as a means of substantiating biological similarity [4-8]. Some of these efforts are anchoring such data to key events within adverse outcome pathways (AOPs) [8]. More information about AOPs themselves can be found in associated references [9-11]. Examples of AOP informed Integrated Approaches to Testing and Assessment (IATA) [12] based on read-across have been developed as part of the Organisation for Economic Co-operation and Development (OECD) work programme under the auspices of the Task Force of Hazard Assessment (TFHA)[1] (see http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm for a list of case studies both published (nine as of the time of writing) and under review (four at the time of writing)).

[1] TFHA has since been renamed to the Working Party on Hazard Assessment (WPHA)

Although there is extensive technical guidance available [13-15] which describes the workflow of category/analogue development and associated read-across, many challenges still remain. The consistency in how read-across predictions are derived, the level of evidence required to support a read-across justification, the uncertainty or confidence in the underlying analogue data and the documentation required, have all impeded greater acceptance of read-across for regulatory purposes [16-18]. Many researchers within academia and government agencies, as well as chemical industries that are being regulated, have been actively working in an attempt to clarify the issues and overcome these challenges. For instance, the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) sponsored a task force to characterise the state of the art in read-across [19-20]. The insights described in the task force report [19] formed part of the discussions at an expert workshop organised by the European Chemicals Agency (ECHA) with active support from the European Chemistry Industry Council's Long Range Initiative (Cefic-LRI) ([21]; http://cefic-lri.org/wp-content/uploads/2014/03/ECHA-Cefic-LRI-Read-across-Workshop-Report_171211-FINAL.pdf). ECHA's draft of the Read-Across Assessment Framework (RAAF) (https://echa.europa.eu/news-and-events/events/event-details/-/journal_content/56_INSTANCE_DR2i/title/experts-workshop-on-read-across-assessment-with-active-support-from-cefic-lri), a consistent set of principles for evaluating read-across justifications submitted under REACH, was first presented at this workshop prior to its initial publication in 2015 [22-23]. Elsewhere, the Center for Alternative to Animal Testing (CAAT) initiated a cross stakeholder workgroup including representatives from academia, industry and governmental agencies to facilitate read-across use. A white paper was published [16], and a team of ~30 experts then set about describing good read-

across practice along with reasons related to non-acceptance by ECHA [18]. Case studies to illustrate the extent to which HT/HC screening data could be useful in capturing biological similarity in conjunction with the traditional chemical similarity approaches were also developed [7, 18]. Several of the most recent European Union (EU) research programmes have been focused on moving away from traditional animal testing – the SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing) programme (http://www.seurat-1.eu/) was a particular example that included a significant read-across component [24]. Templates for structuring and reporting of read-across predictions were developed and applied to a set of repeated dose toxicity case studies to compare and contrast traditional read-across as well as the support of read-across justifications through the integration of data from the so-called New Approach Methodologies (NAM) ([8, 25] and references within). Further work on read-across is also being undertaken in the on-going EU programme EU-ToxRisk (see http://www.eu-toxrisk.eu/).

As a result of its widespread use supported by on-going research activities, there have been many efforts to refine and improve the manner in which read-across is performed and documented. There are several different frameworks available (both to undertake and evaluate read-across), described in both regulatory technical guidance and peer-reviewed scientific literature. For an end-user, understanding the similarities and differences of these different frameworks as well as their application or decision contexts represents a steep learning curve in and of itself. In our previous article [26], we described a generalised workflow for analogue and category development and illustrated where a selection of publicly available software tools aligned in this workflow. The intent was to help guide an end-user through the 'minefield' of available tools and provide some context on where these

tools could be most useful depending on the decision of interest. In a similar manner, this article is aimed at providing a clarifying perspective on the different frameworks/workflows that exist to either develop category and analogue approaches or to assess their associated read-across justifications. A harmonised framework is also proposed to help reconcile the common guiding principles and steps of the read-across process which should be helpful in expanding the scope and decision context of the existing frameworks. This harmonised framework is also intended to provide context for how generalised and systematic read-across approaches, taking into consideration new approach methodology (NAM) information, are evolving. Indeed, using NAM data can be an aim in itself, as such grouping of substances could be beneficial in other contexts of assessment not only as a part of a regulatory read-across assessment.

Our earlier article [26] has already described many of the terms of reference associated with read-across, we only briefly highlight these for ease of comprehension. We then describe a handful of the available workflows including whether they are aimed at developing read-across and/or assessing read-across. We summarise the similarities and differences of these different workflows to make explicit how and where these are consistent with each other. We also propose a harmonised hybrid workflow to reconcile the existing frameworks and provide a direction for how systematic read-across is evolving to include where the new approach methodology (NAM) approaches can and are being practically utilised. Figure 1 provides a graphical outline of this article for ease of reading.

2. Terms of reference

In brief, groups of substances with similar human health and/or environmental toxicological properties typically based around some concept or aspect of chemical similarity are known as chemical categories. A category of two substances (a target which is the substance of interest and a source analogue with data to read-across) is referred to as an analogue approach. Read-across is reserved as a term to describe one of the main data gap filling techniques within category or analogue approaches and can be qualitative or quantitative. Other data gap filling techniques include trend analysis and external Quantitative Structure-Activity Relationships (QSARs) [as described in reference 26].

2.1 Considerations before embarking on a read-across

Considerations before embarking on a read-across approach have been discussed at length elsewhere [16, 17, 19, 20, 27]. The types of considerations vary to a degree depending on the decision context; e.g. the objective may be to extract chemical categories from a starting large inventory of substances (termed a 'top-down' approach) or the context might be to start with a specific target substance and identify suitable source analogues (termed a 'bottom up' approach) (described in more detail in [28-29]). In this article, we focus on the 'bottom up' decision context, i.e. where the starting point is considering a specific target substance and building up an analogue or category approach from appropriate source analogues.

Other considerations depend on the degree of uncertainty that can be tolerated in the read-across prediction as well as the magnitude of effort and resources that can or should be brought to bear to address the overall decision i.e. risk-decision/safety assessment. These encompass the number of data gaps and for which endpoints, a plausible hypothesis for grouping substances together and the ease and cost of substantiating that hypothesis, as well

as the legitimate access to sufficient, high quality and relevant data [13, 16-17, 26]. Some of this information may not be known *a priori* but may be elucidated once a more detailed evaluation of the associated data for source analogues is undertaken.

2.2 Sources of regulatory and technical guidance

Category and analogue approaches have been in use for many years. One of the first instances of read-across in a regulatory context was described by Hanway and Evans (2000) of the UK Health and Safety Executive for human health endpoints [30]. Reviews of regulatory use of chemical categories and read-across were prepared as part of the OECD and European Commission's Joint Research Centre (EC JRC) work programmes in preparation to the development of the REACH technical guidance [27-29]. The reviews revealed that technical guidance for categories was first developed by the United States Environmental Protection Agency (US EPA) in support of the US High Production Volume (HPV) Challenge Program in 1998 [31]. The same principles were embedded into the OECD Manual for the Assessment of Chemicals used as part of the OECD HPV programme, specifically, Chapter 3 of the manual provided guidance for grouping of substances and SARs [32]. In the run up to REACH regulation [1], the EC JRC in collaboration with OECD developed technical guidance for category formation and read-across which could serve to satisfy the needs for REACH as well as global regulatory requirements. The OECD guidance for grouping was published in 2007 [13]. The ECHA guidance (which is almost identical to the OECD guidance) was published several months later as "R6 QSARs and Grouping of Chemicals" [15]. In 2012, ECETOC formed a task force to describe the state-of-the art of category approaches in order to provide some practical guiding principles for Industry registrants that would be of help with

their EU REACH dossier preparation. The ECETOC task force report (TR116) was published in late 2012 [19] and an executive summary from it formed one of the background papers for the ECHA expert workshop with active support from Cefic LRI in October 2012 [21]. Given the insights gained from read-across justifications prepared by industry and evaluated by OECD member countries, coupled with developments in Adverse Outcome Pathways (AOPs), the OECD established a new drafting group to revise the 2007 grouping guidance. The revision was published in 2014 [14] and in fact took up a number of the recommendations from the ECETOC report [19] as well as other publications that had been published in the peer reviewed literature since 2007, e.g. [33-34]. Although ECHA have not updated their original guidance, they have published practical guides on the use of read-across (https://echa.europa.eu/documents/6362380/7127661/pg_report_readacross_en.pdf/69860e5b-c669-4a0d-b868-72f5dba5b560), an illustrative case study (https://echa.europa.eu/documents/10162/13628/read_across_example_1_en.pdf), in addition to their read-across assessment framework (RAAF) [22-23].

Aside from the regulatory technical guidance, there are a number of articles in the peer reviewed literature that have described frameworks for how to develop read-across justifications. Although the needs of REACH [1], CLP [2] or the EU Cosmetics Regulation [3] have provided significant momentum, these other frameworks are sufficiently generic to serve other risk and/or safety assessment purposes. Notable examples include the framework for using structural reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogues for (structure-activity relationship (SAR) based toxicological assessments by Wu et al [33]. This was followed up by a set of case studies

published by Blackburn et al [34] to illustrate the practical application of the approach. Wang et al [35] published a tiered surrogate approach for use in human health risk assessment based on the insights and experiences used within the EPA's process for deriving screening-level provisional peer reviewed toxicity values (PPRTVs) for data-poor chemicals of interest to the EPA's Superfund program [https://hhpprtv.ornl.gov/]. Patlewicz et al [20] summarised the key insights from the ECETOC report to articulate the general considerations for using category approaches. They also summarised the main learnings from the ECHA-CEFIC LRI workshop [21] and updated the earlier paper [20] based on industries' own experiences under REACH [17]. The CAAT Working Group's recommendations and best practices were summarised by Ball et al [18] who additionally described some of the main shortcomings in the read-across justifications submitted to ECHA based on a review of the ECHA published decisions.

In summary, there are several frameworks focused on the development of read-across justifications. These include the ECHA guidance [15], the OECD guidance [14], Wu et al [33], Wang et al [35] and Patlewicz et al [20]. These are compared and contrasted to better articulate their similarities and differences.

3. Frameworks for the development of read-across

3.1 Category/Analogue approaches as described by OECD and ECHA

As indicated earlier, the guidance for analogue and category development was jointly developed by OECD and EU JRC to address REACH [1] and other regulatory frameworks at the same time. These include chapters that specifically describe a stepwise procedure to the

analogue and category approach. There are very minor variations in these stepwise procedures between the 2 documents, which we highlight below.

3.1.1 ECHA (2008) Stepwise procedure to the analogue approach

The first step in the ECHA approach [15] is to search for potential sources analogues for the target substance – see Figure 2. The next step is to gather data for the source analogues and evaluate their fitness for purpose which can be done, in part, by constructing a data matrix. The adequacy of the read-across is then assessed; if this is sufficient, the data gap is filled and the prediction is documented. If the read-across is not adequate – either additional analogues are searched for, or new experimental data are generated.

3.1.2 ECHA (2008) stepwise procedure to category development

ECHA's [15] category approach is very similar to the analogue approach (see Figure 3), save for an additional step that calls for a check on whether an existing regulatory category might already be available. If this is the case, this could be updated with new data and the category re-assessed to determine that it still is valid for its intended purpose. If no existing category is available, the procedure calls for a hypothesis to be developed to aid in the identification of source analogues. One other notable difference with the category approach is that an adequate category might not be possible to construct.

3.1.3 OECD Stepwise approach to an analogue approach

The procedure outlined in the OECD grouping guidance [14] differs only marginally with the ECHA approach in terms of the actual workflow diagram (see Figure 4). A Step 0 has been introduced in the approach to consider whether a target substance is a member of an existing

category. Step 6, which involves documenting the read-across, specifically calls for the justification to be documented.

3.1.4 OECD stepwise approach to category development

The ECHA and OECD workflows are more or less identical save for one item – the ECHA guidance calls for a re-assessment of an existing category if new data are available. This step is not explicitly captured in the OECD stepwise approach though is described in the associated text.

Although the workflows are to all intents and purposes the same between the OECD [14] and ECHA [15], the main significant difference is the extent to which the OECD guidance describes how other types of data can be used in the construction of the data matrix such as mechanistic data anchored to key events within an AOP. Within Step 3 of the category approach "Evaluate available data for adequacy", practical steps on how to group chemicals using an AOP are offered. Specific references are made to the AOP for skin sensitisation that was published by the OECD in 2012 [36].

It is important to consider these workflows in light of prior guidance. Specifically, the OECD [13] and ECHA [15] technical guidance, represented the first time a connection between QSAR approaches and categories was made. It is now well established that the underlying basis for both approaches is the same, but at the time these guidance documents were first drafted, this represented a major step forward in terms of expressly articulating their connection. The guidance document also sought to clearly define key terms of reference to avoid ambiguity, for example 'read across' (by convention written as read-across) had been used synonymously with what is now termed the analogue approach rather than strictly as

a data gap filling technique. The delineation of other data gap filling techniques was also expanded and extended in light of the advances made in QSAR development and its application for regulatory purposes beyond screening or prioritisation. The manner in which key information of the approach should be documented in a consistent format was also novel. The formats, so named reporting formats, provided a much more structured manner for performing and documenting the justification and rationale for the category or analogue approach. The concept of the reporting format was largely driven by similar formats derived to document QSAR models and their associated predictions [15]. The way in which the approaches were structured was subsequently helpful in the development of the associated OECD QSAR Toolbox, a tool aimed at the development, evaluation, justification and documentation of chemical categories [37].

One of the shortcomings in the OECD and ECHA guidance (which in many respects prompted the OECD revision) was that, as much as it described the steps of how to derive a category or analogue approach, it lacked sufficient information or guidance for how the approach should be documented and the level of justification that should be merited – for example how to assess the adequacy of the approach in terms of its uncertainties. At the time of development, there were no examples to draw from aside from the experiences of the OECD and EPA HPV programmes [38] and the insights derived from the use of category and QSAR approaches as they had been applied within the EU Technical Committees for Existing Chemicals (EU TCNES) or Classification and Labelling (EU TC C&L). These insights and experiences were drawn upon extensively in the initial drafting of the guidance and were compiled as a compendium of case studies [27] that were also published and cited in the guidance itself

[13,15]. The issue of adequacy or sufficiency of the justification is discussed in a later section (section 4) where we describe frameworks intended to assess read-across.

3.2 Framework by Wu et al (2010)

Wu et al [33] presented the "systematic expert-driven process" that the Procter and Gamble Company uses to evaluate analogues for read-across in SAR-based toxicological assessments. One of the drivers for this publication was in response to some of the shortcomings the authors themselves had noted of the original OECD [13] and ECHA [15] guidance which they felt was insufficient for assessing the adequacy of source analogues i.e. the analogue evaluation step. The article describes a flowchart (see Figure 5) to outline the overall source analogue identification and evaluation process. It also describes the ranking of source analogues in more detail with respect to different similarity contexts. In brief, the first step of the overall workflow relies upon a chemistry expert(s) to review the target substance in conjunction with other tools/resources (such as expert systems, structure searchable databases, literature etc.) to devise the optimal search strategy that will take into account key functional groups and features and how these might impact the physicochemical, reactivity or metabolic profile. Based on the search performed, the results are then filtered to retrieve only those source analogues with relevant toxicity data.

An assessment of the source analogues (see Figure 6) is then performed to evaluate their suitability on the basis of their structural, physicochemical characteristics, reactivity and

metabolism. For the initial search, a Tanimoto[2] threshold of 0.75 is used as a default to limit the number of structurally similar analogues retrieved. This threshold may be modified depending on what is known or can be inferred about how the toxicity is driven (e.g. toxicity driven by a specific reaction centre). Evaluating the structural differences between the source analogues and the target is then performed to appreciate whether any of those differences would lead to a significant difference in the reactivity and toxicity anticipated. The practical manner in which this evaluation is performed is to evaluate: (1) the commonality of structural alerts (such as those contained within expert systems such as Derek Nexus (Lhasa Ltd)); (2) the commonality of key functional groups (that would be critical for driving the reactivity and sites for metabolism); (3) the commonality in position of double bonds and, (4) the effects of additional functional groups. This multifaceted evaluation would be followed by an assessment of the similarity of the physicochemical characteristics of the source analogues relative to the target. Parameters include the log Kow (the log of the octanol-water partition coefficient), molecular weight (MW), water solubility as well as log D (the distribution coefficient) and pKa (the acid dissociation coefficient), all of which 'model' likely bioavailability. Significant differences in the physicochemical parameters of the source analogues relative to the target substance are considered helpful in identifying or rationalising any differences observed in *in vitro* or *in vivo* toxicity studies [33]. The last consideration is the similarity of the metabolic profile between the target and source substance. The aspects considered are whether there exists the potential of the

[2] Note there are many aspects to take into account when using Tanimoto as a similarity threshold including the manner in which the source analogues and their inventory are characterised to enable rapid searching, as well as what an appropriate cut off might be. Some of these have been discussed elsewhere – see Willett et al [39]

source analogue to metabolise to the target or vice versa, whether the metabolism of source analogue and target diverge to different pathways or converge to similar pathways, both of which will have an impact on the toxicity observed.

The considerations articulated are intended to help in performing a comprehensive evaluation of the suitability of the source analogues identified for a toxicity assessment.

3.3 Tiered surrogate approach by Wang et al (2012)

Wang et al [35] describe a tiered surrogate approach based on identifying three main types of potential surrogates to ultimately select the "best" surrogate for use in a quantitative risk assessment. The decision tree/workflow is shown in Figure 7.

The first step involves understanding what is known about the target substance in terms of its available data and any inferences that can be made about its reactivity, metabolism and toxicity. If no adequate repeated dose toxicity information is available for the target substance that allows for quantitative risk assessment, a search is made to identify surrogates based on structural, metabolic and toxicity considerations. For the first surrogate type – structural considerations include identifying structurally similar substances using the Tanimoto similarity index with a suitable cut off. Assessment of structurally similar surrogates also considers similarity in key functional groups and reactivity making use of structural alerts. Metabolic surrogates are the second type and include metabolic precursors, metabolites and (bio)degradation products/precursors. Literature data or toxicokinetic testing will inform this surrogate type. Potential metabolic surrogates and the target substance are expected to have a similar toxicological profile or mode of action that may result in the same ultimate toxicity at the same target organ or tissue site. The target

organ/tissue would also be noted at this stage especially if the target organ/tissue is known for the target substance. The third type of surrogate is toxicity-like – here reference is made to similar dose-response curves based on a toxicity equivalence factor (TEF) or relative potency factor (RPF). Other validated *in vitro* dose response data could also be considered to establish a potency ranking. Additionally, similarity considerations with respect to common target organs, toxic effects, mode of action (MoA) and group membership (as in well-defined chemical classes/mixtures) are taken into account when identifying this surrogate type. The tiered surrogate approach is reliant on the surrogates identified being associated with established reference /toxicity values from regulatory agencies in order to compile a pool of data-rich surrogate candidates with good quality repeated dose information, facilitating the read-across process. At this stage, physicochemical parameters are collected and added to the pool of information. For the target substance, an assessment, as far as possible, of the likely target organs/tissues is made to gauge and help compare the suitability of any of the surrogates identified. A weight of evidence (WOE) approach is then used to rank the surrogates on the basis of structural, metabolic and toxicity similarity considerations in order to identify the "best" surrogate. During this approach, emphasis is given to biological similarity (toxicity/toxicokinetic) over structural similarity. The most biologically relevant surrogate with the highest similarity score and/or most health-protective toxicity value is selected, adopting its point of departure (POD) for quantitative risk assessment of the target compound.

3.4 Category approaches and read-across considerations by Patlewicz et al (2013)

Patlewicz et al [20] structured a workflow for category/analogue development around the category/analogue reporting format that is described in the OECD [14] and ECHA [15] guidance. Although a workflow was not structured as a graphic, a series of steps were re-articulated as steps of read-across development in a subsequent article [17] – see Figure 8.

In brief, the first step is to understand the decision context under consideration to determine the magnitude of resources that need to be applied to address the question as well as the amount of uncertainty that can be tolerated in the outcome. The next step seeks to identify the number and type of data gaps for the target substance. These will inform the way in which source analogues are identified (the overarching rationale[3]). For example, a custom search for an endpoint that is mechanistically well understood such as skin sensitisation will differ than an analogue approach based on a metabolic precursor similarity context for systemic toxicity endpoints. The next steps involve searching for and evaluating the source analogues identified. This step is critical to evaluate the validity/suitability of the analogues for the data gaps being filled. The evaluation covers assessment of the similarity in the physicochemical profile of the analogues and removal of significant outliers, assessing the metabolic pathway similarity of the analogues to evaluate commonality in key functional groups including reactivity as encoded in structural alerts. The next step involves filling the data gaps either by qualitative or quantitative read-across or through a trend analysis should a clear trend between a specific parameter and the endpoint of concern be apparent. Finally, an assessment of the uncertainty associated with the prediction is made – usually on the

[3] The term "Overarching hypothesis/rationale" was used in Patlewicz et al [17] to denote the basis for grouping chemicals together – metabolic pathway, change in chain length etc. It was coined to describe the different contexts by which categories could be defined per the OECD technical guidance [14]. Here the intent is to provide the basis for how analogues might be practically identified which can differ if the data gap is specific to one endpoint vs many endpoints.

basis of the uncertainty relating to the similarity rationale itself as well as the uncertainty in the underlying data for the source analogues identified. Pradeep et al [40] provides a practical illustration of addressing uncertainty. Practical suggestions for the types of tools and resources that could be applied to address these steps are also described in the original manuscript [20].

Comparing the different read-across development frameworks illustrates much commonality in their structure, purpose and content even if the stated purpose is different (Table 1).

Table 1: Comparison of selected read-across development frameworks

Framework	**ECHA [15]**	**OECD [14]**	**Wu et al [33]**	**Wang et al [35]**	**Patlewicz et al [20]**
Context	REACH	International regulatory purposes	Product Stewardship	Quantitative risk assessment	Regulatory purposes & Product stewardship
Approach	Analogue/Category	Analogue/Category	Analogue	Analogue	Analogue/Category
Method	Aim is to fill an endpoint specific study. Focused on structural similarity as a starting point	A generalisation of the ECHA approach	Systematic stepwise evaluation of analogue suitability based on structure, reactivity,	Approach is based on a weight of evidence (WOE) assessment from structure, ADME[4] and toxicity	Approach is aimed to identify source analogues that can be used to address as many endpoints as appropriate, even

[4] ADME = Absorption, Distribution, Metabolism, Excretion

	Approach is more hypothesis driven		physicochemical and metabolism	considerations. "Best" surrogate is selected from a set of candidates based on most similar and/or most health-protective toxicity value	though the read-across prediction itself is justified on an endpoint per endpoint basis and some source analogues might be excluded from the prediction itself if they are not appropriate for specific endpoints of interest e.g. metabolic analogue excluded from assessment of local

					endpoints such as irritation. Stepwise approach considering general (physicochemical, reactivity, metabolism) and endpoint specific considerations
Terms of reference	Target/Source Analogue	Target/Source Analogue	Substance of interest/Analogue	Chemical of Concern/Surrogate	Target/Source Analogue
Scope	Endpoint specific	Endpoint specific	Most sensitive/relevant endpoint – focused on repeated dose	Most sensitive/relevant endpoint – focused on repeated dose	Most sensitive/relevant endpoint – focused

			and developmental toxicity endpoints; quantitative risk assessment	toxicity endpoints; quantitative risk assessment. Surrogates must be associated with repeated dose toxicity data that has been published as reference/toxicity values.	on repeated dose toxicity endpoints
Use case	Bottom up	Bottom up	Bottom up	Bottom up	Bottom up

There is much similarity in the steps taken and the types of considerations brought to bear in these different read-across workflows. For example, all the selected frameworks propose an initial profiling of the target substance. This involves an analysis of the available information and data gaps that will direct subsequent steps in the read-across workflow such as the analogue search strategy and can therefore help inform the overall appropriateness of the read-across approach. The similarities and differences have been summarised and tabulated in Table 2 using the workflow described in Patlewicz et al [26] as an anchoring framework for convenience.

Table 2: Similarities and differences of selected read-across development frameworks as aligned to the framework described in Patlewicz et al [26]

Framework	**ECHA [15]**	**OECD [14]**	**Wu et al [33]**	**Wang et al [35]**	**Patlewicz et al [20]**
Decision context	REACH	International regulatory purposes	Product Stewardship	Quantitative risk assessment	Regulatory purposes & Product stewardship
Data gap analysis	Not specifically captured as a step in the workflow but implicit based on the accompanying	Not specifically captured as a step in the workflow but implicit based on the accompanying	Not specifically captured as a step in the workflow but discussed in the accompanying text	Not specifically captured as a step in the workflow but discussed in the accompanying text	Stated upfront in the workflow

	text in the guidance	text in the guidance			
Overarching rationale	Includes consideration of existing categories. Captured in the reporting format	Includes an additional step to check whether chemical is a member of an existing category. Captured in the reporting format	Not explicitly referred to as an overarching rationale but captured as part of the initial profiling of the chemical to determine the type of custom search to identify analogues	Not explicitly referred to as an overarching rationale, however the proposed WOE approach considers three primary similarity rationales (i.e. 1) common biological response, toxic effect or MoA, 2) common metabolite/precursor, 3) a TEF or RPF,	Explicitly outlined as a way of factoring in the similarity rationales as a means of structuring and informing the type of analogue search to undertake

				which serve as potential hypotheses for facilitating the selection of the best surrogate.	
Analogue identification	Heavily focused on structure search approaches to identify structurally similar analogues	Heavily focused on structure search approaches to identify structurally similar analogues though reference is	Custom search that takes into account expected metabolism of the target coupled with consideration of key functional groups,	Informed search taking into account profile of chemical of concern to ensure that key functional groups, structural alerts or known/presumed metabolism is considered for the	Unless decision context and data gap analysis inform the specificity of the endpoint, search itself is based on structural similarity

		made to the OECD Toolbox to emphasise the ways in which analogue identification could be influenced based on the endpoint under consideration	structural alerts encoding reactivity although search still relies on using structure searching approaches and setting a threshold for the similarity to limit the number of analogues identified	identification of surrogate candidates. Search still relies on using structure searching approaches	

Analogue evaluation	Largely focused on the underlying data on the analogues and constructing this into a data matrix to facilitate an assessment of consistency and concordance. Should include collection of physicochemical information.	Largely focused on the underlying data on the analogues and constructing this into a data matrix to facilitate an assessment of consistency and concordance. Should include collection of physicochemical information.	A decision tree explicitly lays out the specific questions to assess the suitability of the analogues identified with respect to their physicochemical profile, metabolic profile, structural feature differences and	Surrogates identified via structural similarity searches are filtered based on availability of reference/toxicity values for repeated dose exposure Surrogate data is collated and evaluated based on reactivity, physicochemical, metabolic and toxicity (similarity in mode of action (MoA) or	General considerations of physicochemical characteristics, metabolic similarity, structural dissimilarity and reactivity potential are evaluated. A second step considers endpoint specific considerations that will impact the data gap filling itself

			reactivity profile.	target organ/tissue) considerations	
Data gap filling	Evaluation of analogues is more closely captured in the step "assess the adequacy of the analogue approach and fill the data gap" Aspects such as MoA information, (Q)SAR	Evaluation of analogues is more closely captured in the step "assess the adequacy of the analogue approach and fill the data gap". Aspects such as MoA information,	Not described in this framework – discussed in more detail in [41]; see section on frameworks for read-across evaluation	The surrogate POD (e.g. No observed adverse effect level (NOAEL), Lowest observed adverse effect level (LOAEL), benchmark dose level (BMDL)) used to derive the reference/toxicity value can be carried forward into a quantitative risk	Qualitative/Quantitative read-across (often the most sensitive value); Trend analysis, External QSARs.

	information, key functional groups and biochemical process are considered to evaluate the utility of the source analogues and their robustness in driving the read-across prediction	(Q)SAR information, key functional groups and biochemical process are considered to evaluate the utility of the source analogues and their robustness in driving the read-across prediction		assessment. POD value from the "best" surrogate which is most similar and/or most health-protective	

Uncertainty assessment	Not explicitly described	Not explicitly described	An uncertainty scale provides some context of how the analogues might then be carried forward into the toxicity assessment. The scale aligns the different types of analogues relative to the uncertainty in the type of read-across that	Not explicitly described	Explicit in the workflow but is discussed herein in the section on read-across evaluation frameworks

			might be conducted based on the consistency and correspondence of the analogue data and the number of source analogues under consideration		

4. Frameworks for Assessing Read-across

There have been many efforts to explore how read-across justifications can be critically evaluated in order for their uncertainties to be explicitly identified and to help focus what practical strategies can be used to reduce those uncertainties. In this manuscript, articles and guidance that are more focused on evaluating read-across are denoted 'read-across assessment frameworks'. In many respects, this designation is arbitrary as the factors driving uncertainty are indirect guidance for read-across development itself. Nonetheless, the distinction is helpful to discriminate guidance that was specifically tailored to describe the steps to follow in a workflow to develop a category/analogue approach versus considerations to bring to bear when attempting to justify and strengthen a read-across prediction. Blackburn and Stuard [41] followed up from work by Wu et al [33] and Blackburn et al [34] to propose an assessment framework aimed at promoting greater consistency in read-across predictions for repeated dose and developmental toxicity endpoints [38]. Their article combined the SAR assessment suitability ranking described in Wu et al [33] with the sources of uncertainty in the source analogues themselves in order to provide qualitative confidence scores for any read-across performed. The framework comprises three parts. Part I reiterates the analogue suitability assessment described in Wu et al [33]. Part II comprises a series of questions to probe the consistency and quality of the data underpinning the source analogues identified. The types of questions address the number of studies addressing the endpoint for any of the source analogues; the quality of the studies for those source analogues; evidence in the entire source analogue set for a specific toxic effect for the endpoint; the concordance in endpoint specific effects and/or PODs across analogues

suggesting similar potency for the critical effect in the endpoint; whether the source analogues demonstrate low or no toxicity as well as whether the most sensitive POD was being used as the read-across threshold. Part III considers the consistency between analogue and target data across all endpoints. The question considers the concordance across the data in terms of the adverse effects indicating that the analogues share the same targets of toxicity. The overall framework considers the sources of uncertainty, namely the number of analogues contributing data, the robustness of the analogue dataset, the concordance of effects and potency, severity of critical effects in order to assign an overall uncertainty category and an associated default uncertainty factor. Four categories representing graded degrees of uncertainty were proposed. A low uncertainty category would be assigned an uncertainty factor of 1, whereas a low-moderate uncertainty would drive an uncertainty factor of 3. A moderate uncertainty would drive an uncertainty factor of 10 and finally a high uncertainty would mean that the use of read-across for data gap filling was not recommended.

4.1 Scientific confidence considerations in read-across by Patlewicz et al (2015)

Table 2 in Patlewicz et al [17] articulated very similar sources of uncertainty as described in Blackburn and Stuard [41]. The sources of uncertainty were categorised into two types – uncertainties associated with the underlying data of the source analogues themselves and uncertainties associated with the similarity rationale between the source analogues relative to the target substance. They tabulated strategies to reduce uncertainties dependent on the similarity rationale element – metabolic transformation, unspecific toxicity, specific toxicity or structural dissimilarities.

4.2 Strategy for structuring and reporting read-across by Schultz et al (2015)

Schultz et al [25] outlined a strategy for structuring and reporting a read-across prediction. Four main read-across scenarios (direct acting substances with similar MoA; indirect acting substances with similar MoA; low or no toxicity hence not reactive or no specific MoA; and, structurally similar substances with different MoA) are described which in essence outline the types of overarching rationales by which analogues/categories might be formed. Assessing the validity of those source analogues is carried out with respect to different similarity contexts such as structure, physicochemical characteristics, toxicokinetic profile, abiotic/metabolic profile, toxicophore/structural alerts, mechanistic profile (same molecular initiating event (MIE) linked to an AOP) and *in vivo* toxicological responses. These components form one part of the overall uncertainty assessment – assessing the similarity of the analogues. For each similarity context, an evaluation is made to grade the data uncertainty underpinning each context (whether that data were modelled or experimental in nature) and the strength of evidence supporting that context. The evaluation is facilitated by a template to record each similarity context in turn, the respective gradings scaled from low - high and any comments. The basis for the assessment is driven by information collected in the data matrix structured as part of the typical read-across workflow. The other component of the uncertainty assessment address mechanistic relevance and the completeness of the read-across. These factors here comprise the number of analogues, the absence/presence of toxicity, the quality of the underlying data for the source analogues, the consistency and concordance in the data and potency across the analogues including a consideration whether the temporal and dose response relationship between mechanistically relevant endpoints is consistent. For this uncertainty component, each

factor is weighted in the same grading scale as part of a weight of evidence assessment. The weight of evidence call reflects the overall uncertainty of the read-across itself. The grading scales used for the different parts of the uncertainty assessment are a modification of what Blackburn and Stuard [41] proposed as part of their own uncertainty assessment framework. Instead of 4 grades to the scheme, Schultz et al [25] have focused on 3 grades, the fourth had been a flag to indicate that a read-across was not appropriate because the uncertainty was too high.

4.3 The ECHA Read-across Assessment Framework (RAAF)

The ECHA RAAF [23] is intended to provide a framework and guidance for consistent evaluation of the scientific validity of a read-across. Although aimed at ECHA evaluators, it was published to assist industry registrants in developing their own read-across justifications when attempting to satisfy information required by making use of the adaptations in Annex XI of the REACH regulation [1]. The RAAF is structured to describe a number of different read-across scenarios. For each of these scenarios, a number of scientific considerations can be identified. Assessment elements are aligned with each consideration which carry over into an appraisal to inform decision making. There are 6 scenarios in total – two of these refer to analogue approaches, the other 4 are applicable to category approaches. There are 2 types of scenarios – one where the hypothesis is based on (bio)transformation to a common compound and the other where the hypothesis is based on different compounds having qualitatively similar properties. The only aspect that subcategorises these scenarios further, relates to whether there are quantitative variations in the properties observed across the category members. Hence, for the analogue approach

there are the same 2 read-across hypotheses. For a category approach, there are 4 scenarios to account for the fact that there may be quantitative variations for both read-across hypotheses. For each scenario, there comprises a pre-defined set of assessment elements (AEs) which are intended to cover all the essential scientific considerations that need to be addressed. Each assessment element is then scored from 1-5, where 5 is designated "acceptable with high confidence" through to 1 which is not acceptable. These scores are termed assessment options (AOs). A minimum AO score of 3 – acceptable with sufficient confidence is needed for a read-across to be taken up and used to inform decision making. There are a number of common assessment elements – common since they are considered with any of the 6 scenarios as well as scenario specific elements. For example, a common assessment element might address the reliability and adequacy of the underlying analogue data. In contrast, a scenario specific assessment element might include: common underlying mechanism, exposure to other compounds than to those linked to the prediction as noted for analogue scenario 2 – different compounds have qualitatively similar properties. The outcome of the read-across assessment takes the form of a conclusion whereby the set of all individual AOs obtained for each of the AEs in the applied scenario are considered. Since all AEs are considered essential, all need to have a minimum AO of 3 or higher for the read-across approach to be acceptable. The intention is that delineating the AEs in this manner with scores will facilitate an evaluator or indeed registrant to quickly identify the "weakest link" in the read-across justification and help focus where additional evidence could be generated or collected to strengthen the justification. The type of "additional evidence" noted includes mechanistic information from *in vitro*, *in chemico* studies or *in silico* predictions as well as toxicokinetic data from *in vivo* or *in vitro* studies. The ECHA RAAF [22, 23] was

originally structured to address human health endpoints alone but has since been extended to capture environmental fate properties and environmental hazards as well as UVCBs (Substances of unknown or variable composition, complex reaction products or biological materials) [42].

Table 3 compares and contrasts the selected read-across assessment frameworks. On the whole, all are quite similar to each other. The differences lie in the extent to which they are described and the extent to which the uncertainties once identified and evaluated might be addressed and reduced. The RAAF differs in terms of its level of prescription using the scenarios and the associated assessment elements. However, even these assessment elements bear much similarity to the similarity rationales described in the other frameworks. The only other difference in the RAAF is its intended purpose, a minimum score needs to be reached for any one assessment element in order to determine acceptance. The framework and considerations in Patlewicz et al [17] emphasise the practical means of how similarity rationales might be characterised and what additional data could be generated to reduce residual uncertainty [43]. As described earlier, the assessment frameworks provide strategies for characterising and dealing with potential sources of uncertainty in both the underlying data for the source analogues as well as the similarity considerations pertaining to the read-across itself. Some of the shortcomings of these approaches should still be noted, including the lack of strategies for integrating and evaluating different types of data/evidence (i.e. *in vivo*, versus *in vitro* versus, *in silico*) and their associated properties (i.e. strength, relevance, reliability, etc). Additional efforts are needed that draw from more established and systematic approaches for weighing evidence such as in the case of MoA/AOP-based risk assessment for which there is ample of literature and even technical

guidance documents [43-47] as well as new guidance that is evolving out of work from the OECD IATA case studies [12, 48-49].

Table 3: Comparison of the similarities and differences of selected read-across assessment frameworks

Framework	**ECHA RAAF [23]**	**Blackburn and Stuard [41]**	**Patlewicz et al [17]**	**Schultz et al [25]**
Context	REACH	Product Stewardship	Regulatory purposes & Product stewardship	Regulatory purposes & Product stewardship
Scope	Analogue/Category	Analogue/Category	Analogue/Category	Analogue/Category
Framework	Scenarios addressing analogue (2) and category (4) approaches as described above	Framework addresses 3 aspects: analogue suitability (covered in [33]);	Identifies the sources of uncertainty in relationship to the	Different scenarios are articulated to frame up to 11 different similarity criteria. 8 factors proposed to evaluate

	Each scenario is associated with a number of assessment elements (AE) (both common and scenario specific).	data quality of the analogues; consistency of the data across the analogues and relative to the target	data and similarity context	mechanistic relevance and completeness of the read-across
Data issues	Characterised by common assessment elements	No. of analogues, robustness of data, concordance of effects and potency, severity of critical effect	No. of analogues, robustness of data, concordance of effects and potency, severity of critical effect	No. of analogues, robustness of data, concordance of effects and potency, severity of critical effect which are then assessed as a WOE

Similarity rationales	Characterised by scenario specific assessment elements	Analogue suitability rating as described by Wu et al [33]	Structural, physicochemical, reactivity, metabolic, toxicokinetic, and toxicological similarity (extended based on the considerations outlined and described in [20])	Structural, physicochemical, toxicokinetic, metabolic, structural alerts (reactivity), toxicological, mechanistic plausibility and AOP related events
Grading scale	Each AE is scored by an assessment option (AO) from 1-5. A minimum of 3 must be achieved for	Low – High gradings which are associated with default uncertainty	None – possible strategies to reduce	Low to High but no default quantitative uncertainty factors are proposed

	all AEs for a given scenario for the read-across to be acceptable.	factors 1-10. High uncertainty translates to no read-across possible	uncertainties are proposed	

5. The need for a harmonised hybrid framework for read-across

Given that there are many similarities between the frameworks discussed here, a harmonised framework could be readily proposed which integrates the similarity contexts, and sources of uncertainty from the existing frameworks. It is also worth highlighting one aspect that is missing from all the frameworks. Each of these frameworks is, to all extents, qualitative in nature and rely on expert input to evaluate a specific read-across assessment i.e. a read-across is typically performed on a case-by-case basis. There are only a couple of methodological aspects considered within the frameworks that are, to some extent, quantitative. For instance, the structural similarity search that is often employed by each of these development frameworks often relies upon a Tanimoto similarity (Jaccard coefficient) index as a numeric threshold to limit the number of plausible and pragmatic candidate source analogues. The data quality underpinning source analogues is often assessed using the principles outlined by Klimisch et al [50]. This is the standard used under EU regulatory schemes and results in a score on a scale of 1-4, where 1 is the highest quality and termed "reliable without restriction", 3 is of lowest quality and termed "not reliable" whereas 4 denotes "not assignable". However, there are no objective measures to assess the predictive performance of the read-across derived to demonstrate its general applicability nor any objective quantitative characterisation of the uncertainty associated with the prediction made. With regard to the latter, the confidence associated with a read-across prediction is at most assigned by an expert in terms of scores such as low, moderate or high on the basis of building up a body/weight of evidence. Blackburn and Stuard [41] proposed a grading scheme, which is characterised by specific criteria, but still relies on an expert driven

assessment of each component to determine a final uncertainty score and use of a specific uncertainty or assessment factor. This mirrors aspects of current and traditional risk assessment approaches which apply assessment factors to *in vivo* toxicity studies and use them as a basis to derive a reference value. Schultz et al [25] emphasised the need to build up a body/weight of evidence using data that characterise the different similarity contexts as well as the underlying toxicity data for the target and source analogues. However, the relative contribution or significance of any of these similarity contexts or underlying data in driving the overall confidence in the read-across is not well established in objective terms. In contrast, efforts by Low et al [4], examples highlighted in Zhu et al [7] and work by Shah et al [6] have explored the notion of quantifying read-across in a manner akin to approaches relied upon in QSAR. Specifically, in Shah et al [6], the focus was to systematically evaluate read-across performance and characterise its uncertainty for a large number of substances so that the approach could be generalisable and therefore capable of making inferences for large numbers of chemicals rather than require a manual and expert intervention on a case-by-case basis. Further, recent work (Helman et al in prep; Pradeep et al [40]) has quantified the impact and role that other similarity contexts specifically physicochemical properties may play in driving read-across performance over and above chemical structural information alone.

The harmonised framework proposed here is intended as a hybrid approach. The foundation of the harmonised framework is still anchored to the main steps of the current expert driven process taking into account insights and aspects captured in the other frameworks and refining current shortcomings such as exploiting systematic approaches to weighing evidence or strategies for integrating different types of data/evidence together. At the same

time, it exploits NAM and computational strategies to evolve read-across towards a more systematic, reproducible and objective approach i.e. exploring ways in which some read-across aspects such as other contexts of similarity could be quantified where feasible or appropriate (Figure 9). Here the generic workflow in Patlewicz et al [26] has been used as a foundation to demonstrate the extensibility using the aspects captured in the other frameworks and highlighting where aspects of the read-across approach could be quantified, and where novel types of NAM information could be incorporated (Table 4).

Table 4: Harmonised hybrid framework for read-across incorporating aspects from all existing frameworks and introducing quantification of aspects of read-across where feasible

Step	Practical aspects	Comments
1.Decision context	Gauge the scope of the read-across assessment needed i.e. the level of effort and resources that might be required to undertake the read-across assessment workflow. Determine the type of read-across prediction that might be feasible based on the underlying data (categorical or quantitative in nature) and sufficient for the different risk-decision contexts	Regulatory contexts such as risk assessment in e.g. PPRTV and REACH [1], emergency response assessments. Other contexts might cover product stewardship, screening level assessment
2.Data gap analysis	Determine the number and type of data gaps for the target.	Use this to determine whether a data gap filling strategy might be practically best addressed by other techniques. This might be appropriate for endpoints such as physicochemical properties, various ecotoxicity or

		environmental fate properties for which QSARs are typically better established. For a single data gap such as for an endpoint which is reasonably well understood mechanistically or indeed has an associated AOP established, defined approaches might be better applied for data gap filling. Examples include oestrogenicity [47] where a battery of specific HTS assays are integrated into a prediction model or the defined approaches that have been developed to integrate multiple key event information together in a systematic objective manner for the skin sensitisation endpoint [48-49] For other endpoints or for a larger number of data gaps, continuing to step 3 of the workflow may be more appropriate.
3.Overarching similarity rationale	Perform an initial profiling of the target substance to explore the types of effects	This could result in establishing which of the scenarios described in the RAAF [23] or those articulated by Schultz

	observed, whether anything is known about the target organs, mechanisms of action or toxicokinetics to determine whether effects are likely to be driven by the target or its (abiotic, biotic, metabolic) transformation product. This profiling would ideally consider both available empirical data as well as those predicted effects using (Q)SAR to help inform the basis for how analogues should be identified to develop analogue/category approaches.	et al [25] or the OECD grouping guidance itself [14] are most useful in driving the analogue identification step.
4.Analogue identification	Directed by the outcomes of the data gap analysis and overarching rationale steps. This could be as simple as performing a structure search using a similarity metric such as Tanimoto. It could also consider	Scope includes category approaches where assessment of several category members and/or a several endpoints are of interest. Custom searches are described in Wu et al [33] and Wang et al [35].

	other similarity contexts such as physicochemical profiles or structural alerts etc. sequentially (search on one context and subcategorising the outcomes derived) or it could involve a 'search expansion' approach whereby a custom fingerprint of structural characteristics are weighted alongside a custom fingerprint of physicochemical characteristics, presence/absence of structural alerts, structural and topological features from a metabolic profiler, bioactivity fingerprint from HT/HC screening data to search for analogues objectively using a similarity index such as the Jaccard index	Custom searches that might rely on chemical features/properties pertinent for a specific endpoint Objective 'search expansion[5]' approaches as discussed by Helman et al (in prep). It is worth noting that grouping substances on the basis of NAM data such as HT/HC could also be an aim in itself – especially for other types of assessments e.g. efficacy assessments.

[5] In Helman et al (in prep) – approaches for searching for similar analogues on the basis of structural and physicochemical characteristics at the same time were described as a search expansion approach

5.Analogue evaluation	Underlying data and analogue evaluation. Addresses the similarity contexts if not considered explicitly in the analogue identification step. Codify the data quality using schemes such as Klimisch et al [50] and quantity the data variability if multiple studies from the same endpoint are available or apply a global variability factor. Focus on consistency and concordance in effects and potency as well as their temporal and dose response relationship across analogues and endpoints and relative to the target.	This is described in more detail in Schultz et al [25] and to an extent Blackburn and Stuard [41]. Boxplot analysis may aid in comparing variability of data across multiple studies for the same substance. Strategies to understand the role and impact that these different similarity contexts have in driving the endpoint(s) of interest and the performance of the associated read-across are needed. Progress has been made in Helman et al (in prep) for physicochemical similarity. Approaches to evaluate the weight of evidence (WOE) from the different source analogues to codify and to an extent quantify their relative contribution to the read-across required are also needed. Efforts have been made in clarifying an quantifying WOE approaches (see [45-46])

6.Data gap filling	Quantitative or qualitative read-across. Most sensitive value, average, similarity weighted average, other approaches.	This could include both data driven (machine learning, Bayesian inference) and expert driven means to derive a prediction
7.Uncertainty assessment	Apply qualitative grading schemes as described in the RAAF [23], Schultz et al [25] or Blackburn and Stuard [41]. Apply quantitative assessment techniques which could borrow from decision theory relying on Weight of Evidence Bayesian approaches or QSAR approaches to include Area under the Curve (AUC) from a Receiver Operating Characteristic (ROC) curve or root mean square error (RMSE) from regression based prediction models. Both approaches are intended to quantify	Feedback loop to assess whether the uncertainty exceeds what is needed for the overall decision context. Strategies to reduce uncertainties could include targeted generation of new data based on 'value of information' or sensitivity analysis. Type of data that could be generated may include NAM information such as HT data anchored to specific pathways or AOPs.

	performance using appropriate metrics and provide a quantitative measure of confidence for the read-across predictions made	

6. Conclusions

There are many frameworks and workflows described for both the development of read-across and their assessment. Here a comparison has been made of a selection of the most well-established read-across frameworks (development and assessment) that have either been published as technical regulatory guidance or in the peer reviewed literature. We compared and contrasted their similarities and differences and proposed a harmonised hybrid framework that integrates all the different insights together. All the frameworks are (mainly) qualitative in nature and rely on expert judgement in both the development and assessment of a read-across. We highlighted how objective measures of performance assessment, evaluation of data variability and quantitative measures of similarity addressing many different contexts of similarity could be potentially brought to bear in aspects of the read-across framework. Doing so would help bridge the continuum between QSAR and read-across and thus promote consistent reproducible predictions that can be fit for different decision contexts and purposes.

Disclaimer *The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.*

References

[1] European Commission (EC), Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off. J. Eur. Union L396/1 (2006).

[2] EC – European Commission, Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Off. J. Eur. Union L353/1 (2008). Commission of the European Communities.

[3] EC- European Commission, Regulation (EC) No 1223/2009 of the European Parliament and the Council of 30 November 2009 on cosmetic products. Off. J. Eur. Union L342 (2009), 59-209.

[4] Y. Low, A. Sedykh, D. Fourches, A. Golbraikh, M. Whelan, I. Rusyn, A. Tropsha, Integrative chemical-biological read-across approach for chemical hazard classification, Chem. Res. Toxicol. 26 (2013) 1199-1208.

[5] P. Pradeep, R.J. Povinelli, S.J. Merrill, S. Bozdag, D.S. Sem, Novel Uses of In Vitro Data to Develop Quantitative Biological Activity Relationship Models for in Vivo Carcinogenicity Prediction, Mol. Inf. 34 (2015) 236–245. https://doi.org/10.1002/minf.201400168

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

[7] H. Zhu, M. Bouhifd, E. Donley, L. Egnash, N. Kleinstreuer, E.D. Kroese, Z. Liu, T. Luechtefeld, J. Palmer, D. Pamies, J. Shen, V. Strauss, S. Wu, T. Hartung, 2016. Supporting read-across using biological data, ALTEX 33 (2016) 167-182. https://doi.org/10.14573/altex.1601252.

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

[9] D.L. Villeneuve, D. Crump, N. Garcia-Reyero, M. Hecker, T.H. Hutchinson, C.A. LaLone, B. Landesmann, T. Lettieri, S. Munn, M. Nepelska, M.A. Ottinger, L. Vergauwen, M. Whelan, 2014. Adverse outcome pathway (AOP) development I: strategies and principles, Toxicol. Sci. 142 (2014) 312-320. https://doi.org/10.1093/toxsci/kfu199.

[10] S.W. Edwards, Y.M. Tan, D.L. Villeneuve, M.E. Meek, C.A. McQueen, Adverse Outcome Pathways-Organizing Toxicological Information to Improve Decision Making, J. Pharmacol. Exp. Ther. 356 (2016) 170-181. https://doi.org/10.1124/jpet.115.228239.

[11] C. Wittwehr, H. Aladjov, G. Ankley, H.J. Byrne, J. de Knecht, E. Heinzle, G. Klambauer, B. Landesmann, M. Luijten, C. MacKay, G. Maxwell, M.E. Meek, A. Paini, E. Perkins, T. Sobanski, D. Villeneuve, K.M. Waters, M. Whelan, How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology, Toxicol. Sci. 155 (2017) 326-336. https://doi: 10.1093/toxsci/kfw207.

[12] OECD, Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA) Series on Testing and Assessment No. 260. Organisation for Economic Co-operation and Development, Paris, France, 2017.

[13] OECD, Guidance on grouping of chemicals. OECD Series on Testing and Assessment No. 80. Organisation for Economic Co-operation and Development, Paris, France, 2007.

[14] OECD, Guidance on grouping of chemicals. OECD Series on Testing and Assessment No. 194. Organisation for Economic Co-operation and Development, Paris, France, 2014.

[15] ECHA, Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals, 2008. http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf (accessed 9 April 2018).

[16] G. Patlewicz, N. Ball, R.A. Becker, E.D. Booth, M.T.D.Cronin, D. Kroese, D. Steup, B. Van Ravenzwaay, T. Hartung, Food for thought….. Read-across approaches - misconceptions, promises and challenges ahead, ALTEX 31 (2014) 387-396.

[17] G. Patlewicz, N. Ball, P. Boogaard, R.A. Becker, B. Hubesch, Building scientific confidence in the development and evaluation of read-across, Regul. Toxicol. Pharmacol. 72 (2015) 117-133.

[18] N. Ball, M.T. Cronin, J. Shen, K. Blackburn, E.D. Booth, M. Bouhifd, E. Donley, L. Egnash, C. Hastings, D.R. Juberg, A. Kleensang, N. Kleinstreuer, E.D. Kroese, A.C. Lee, T. Luechtefeld, A. Maertens, S. Marty, J.M. Naciff, J. Palmer, D. Pamies, M. Penman, A.N. Richarz, D.P. Russo, S.B. Stuard, G. Patlewicz, B. van Ravenzwaay, S. Wu, H. Zhu, T. Hartung, Toward Good Read-

Across Practice (GRAP) guidance, ALTEX 33 (2016) 149-166. https://doi.org/10.14573/altex.1601251.

[19] ECETOC, Technical Report 116 Category approaches, read-across, (Q)SAR. http://www.ecetoc.org/technical-reports, 2012 (accessed 9 April 2018)

[20] G. Patlewicz, N. Ball, E.D. Booth, E. Hulzebos, E. Zvinavashe, C. Hennes, Use of category approaches, read-across and (Q)SAR: general considerations, Regul. Toxicol. Pharmacol. 67 (2013) 1-12. https://doi.org/10.1016/j.yrtph.2013.06.002.

[21] G. Patlewicz, D.W. Roberts, A. Aptula, K. Blackburn, B. Hubesch, Workshop: use of "read-across" for chemical safety assessment under REACH, Regul. Toxicol. Pharmacol. 65 (2013) 226-228. https://doi.org/10.1016/j.yrtph.2012.12.004.

[22] ECHA, Read-across Assessment Framework (RAAF). ECHA-15-R-07-EN, 2015.

[23] ECHA, Read-Across Assessment Framework (RAAF) ECHA-17-R-01-EN, 2017.

[24] E. Berggren, P. Amcoff, R. Benigni, K. Blackburn, E. Carney, M. Cronin, H. Deluyker, F. Gautier, R.S. Judson, G.E. Kass, D. Keller, D. Knight, W. Lilienblum, C. Mahony, I. Rusyn, T. Schultz, M. Schwarz, G. Schüürmann, A. White, J. Burton, A.M. Lostia, S. Munn, A. Worth, 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 (2015) 1232-1240. https://doi.org/10.1289/ehp.1409342.

[25] T.W. Schultz, P. Amcoff, E. Berggren, F. Gautier, M. Kalric, D.J. Knight, C. Mahony, M. Schwarz, A. White, M.T.D. Cronin, A strategy for structuring and reporting a read-across prediction of toxicity, Regul. Toxicol. Pharmacol. 72 (2015) 586-601.

[26] G. Patlewicz, G. Helman, P. Pradeep, I. Shah, Navigating through the minefield of read-across tools. A review of in silico tools for grouping, Computational Toxicology 3 (2017) 1-18.

[27] A.P. Worth, G. Patlewicz, A compendium of case studies that helped shape the REACH Guidance on Chemical Categories and Read-across, EUR 22481 EN, 2007. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/compendium-case-studies-helped-shape-reach-guidance-chemical-categories-and-read-across (accessed 9 April 2018).

[28] G. Patlewicz, A. Gallegos Saliner, M. Pavan, A. Worth, R. Benigni, A. Aptula, A. Bassan, C. Bossa, A. Falk-Filipsson, V. Gillet, N. Jeliazkova, A. McDougal, J. Mestres, I. Munro, T. Netzeva, B. Safford, B. Simon-Hettich, I. Tsakovska, M. Wallen, C. Yang, Chemical Similarity and Threshold of Toxicological Concern (TTC) Approaches: Report of an ECB Workshop, EUR 22657 EN, 2005 https://pdfs.semanticscholar.org/7d68/8ba443aacfd1601e49f1452f3b1328117134.pdf (accessed 9 April 2018)

[29] G Patlewicz, Chemical Categories and Read-across, EUR 21898EN, 2005 https://core.ac.uk/download/pdf/38613756.pdf (accessed 9 April 2018)

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

[31] US EPA, Data collection and development on high production volume (HPV) chemicals, Fed. Reg. 65 (248) (2000) 81686-81698. https://www.gpo.gov/fdsys/pkg/FR-2000-12-26/pdf/00-32498.pdf (accessed 6 April 2018)

[32] OECD, Chapter 3: Data Evaluation in the Manual for the Assessment of Chemicals, 2000. http://www.oecd.org/chemicalsafety/risk-assessment/manualfortheassessmentofchemicals.htm (accessed 6 April 2018)

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

[34] K. Blackburn, D. Bjerke, G. Daston, S. Felter, C. Mahony, J. Naciff, S. Robison, S. Wu, Case studies to test: A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments, Regul. Toxicol. Pharmacol. 60 (2011) 120-135. https:// doi.org/10.1016/j.yrtph.2011.03.002.

[35] N.C. Wang, Q. Jay Zhao, S.C. Wesselkamper, J.C. Lambert, D. Petersen, J.K. Hess-Wilson, Application of computational toxicological approaches in human health risk assessment. I. A tiered surrogate approach, Regul. Toxicol. Pharmacol. 63 (2012) 10-19. https://doi.org/10.1016/j.yrtph.2012.02.006.

[36] OECD, The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins Part 1: Scientific Evidence. Series on Testing and Assessment No. 168 ENV/JM/MONO(2012)10/PART1, 2012.

[37] S.D. Dimitrov, R. Diderich, T. Sobanski, T.S. Pavlov, G.V. Chankov, A.S. Chapkanov, Y.H. Karakolev, S.G. Temelkov, R.A. Vasilev, K.D. Gerova, C.D. Kuseva, N.D. Todorova, A.M.

Mehmed, M. Rasenberg, O.G. Mekenyan, QSAR Toolbox - workflow and major functionalities, SAR QSAR Environ. Res. 19 (2016) 1-17.

[38] US EPA, Status and Future Directions of the High Production Volume Challenge Program. https://nepis.epa.gov/Exe/tiff2png.cgi/P1004QXK.PNG?-r+75+-g+7+D%3A%5CZYFILES%5CINDEX%20DATA%5C00THRU05%5CTIFF%5C00001370%5CP1004QXK.TIF, 2004 (access 6 April 2018)

[39] P. Willett, J. Barnard, G. Downs, Chemical similarity searching, J. Chem. Inf. Comput. Sci. 38 (1998) 983–996.

[40] P. Pradeep, K. Mansouri, G. Patlewicz, R.S. Judson, A systematic evaluation of analogs and automated read-across prediction of estrogenicity: A case study using hindered phenols, Computational Toxicology 4 (2017) 22-30.

[41] K. Blackburn, S.B. Stuard, A framework to facilitate consistent characterization of read across uncertainty, Regul. Toxicol. Pharmacol. 68 (2014) 353-362.

[42] ECHA, Read-Across Assessment Framework (RAAF) - considerations on multi-constituent substances and UVCBs, ECHA-17-R-04-EN, 2017.

[43] US EPA, Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F, 2005. https://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf (accessed 6 April 2018)

[44] ECHA, Guidance on information requirements and chemical safety assessment Chapter R.4: Evaluation of available information, 2008.

http://echa.europa.eu/documents/10162/13632/information_requirements_r4_en.pdf

(accessed 6 April 2018)

[45] R.A. Becker, G.T. Ankley, S.W. Edwards, S.W. Kennedy, I. Linkov, B. Meek, M. Sachana, H. Segner, B. Van Der Burg, D.L. Villeneuve, H. Watanabe, T.S. Barton-Maclaren, Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence, Regul. Toxicol. Pharmacol. 72 (2015) 514-537. https://doi.org/ 10.1016/j.yrtph.2015.04.004.

[46] R.A. Becker, V. Dellarco, J. Seed, J.M. Kronenberg, B. Meek, J. Foreman, C. Palermo, C. Kirman, I. Linkov, R. Schoeny, M. Dourson, L.H. Pottenger, M.K. Manibusan, Quantitative weight of evidence to assess confidence in potential modes of action, Regul. Toxicol. Pharmacol. 86 (2017) 205-220. https://doi.org/10.1016/j.yrtph.2017.02.017.

[47] E. Julien, A.R. Boobis, S.S. Olin, Ilsi Research Foundation Threshold Working Group. The key events dose–response framework: a cross-disciplinary mode-of-action based approach to examining dose–response and thresholds, Crit. Rev. Food Sci. Nutr. 49 (2009) 682-689. https://doi.org/10.1080/10408390903110692

[48] OECD, Guidance Document for the Use of Adverse Outcome Pathways in Developing IATA. STA No. 260, ENV/JM/MONO(2016)67, Paris, France, 2016.

[49] OECD, OECD Guidance Document on the Reporting of Defined Approaches (DAs) to Be Used within IATA. STA No. 255, ENV/JM/MONO(2016)28, Paris, France, 2016.

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

[51] R.S. Judson, F.M. Magpantay, V. Chickarmane, C. Haskell, N. Tania, J. Taylor, M. Xia, R. Huang, D.M. Rotroff, D.L. Filer, K.A. Houck, M.T. Martin, N. Sipes, A.M. Richard, K. Mansouri, R.W. Setzer, T.B. Knudsen, K.M. Crofton, R.S. Thomas, Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor, Toxicol Sci. 148 (2015) 137-154. https://doi.org/10.1093/toxsci/kfv168.

Figure 1: Graphical guide to the organisation of the manuscript

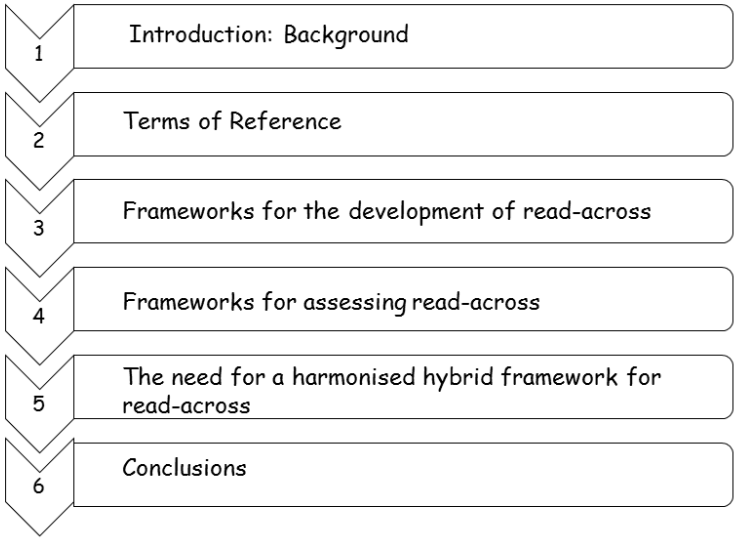

Figure 2: ECHA workflow for the analogue approach (taken from [15])

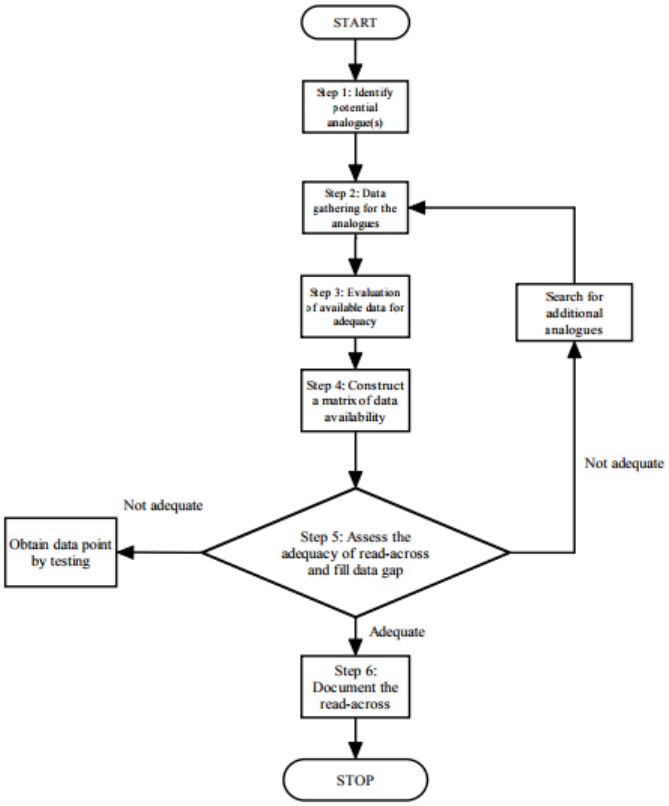

Figure 3: ECHA stepwise workflow for the category approach (taken from [15])

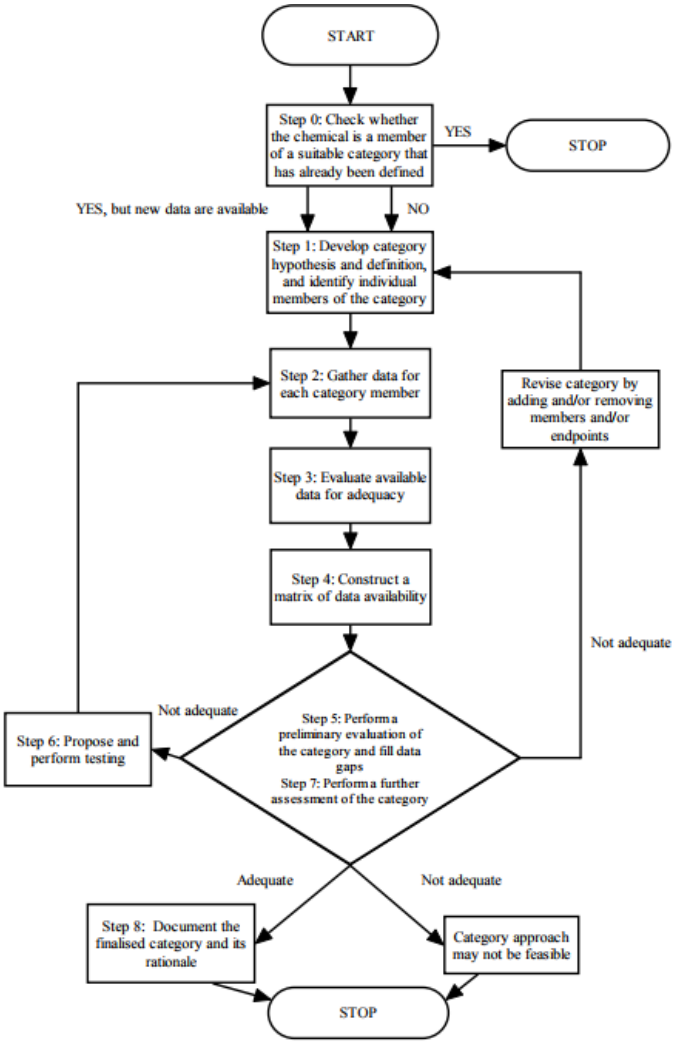

Figure 4: OECD workflows for analogue and category approaches [14]. *Republished with permission of OECD, from Guidance on Grouping of Chemicals, Second Edition, 2014; permission conveyed through Copyright Clearance Center, Inc.*

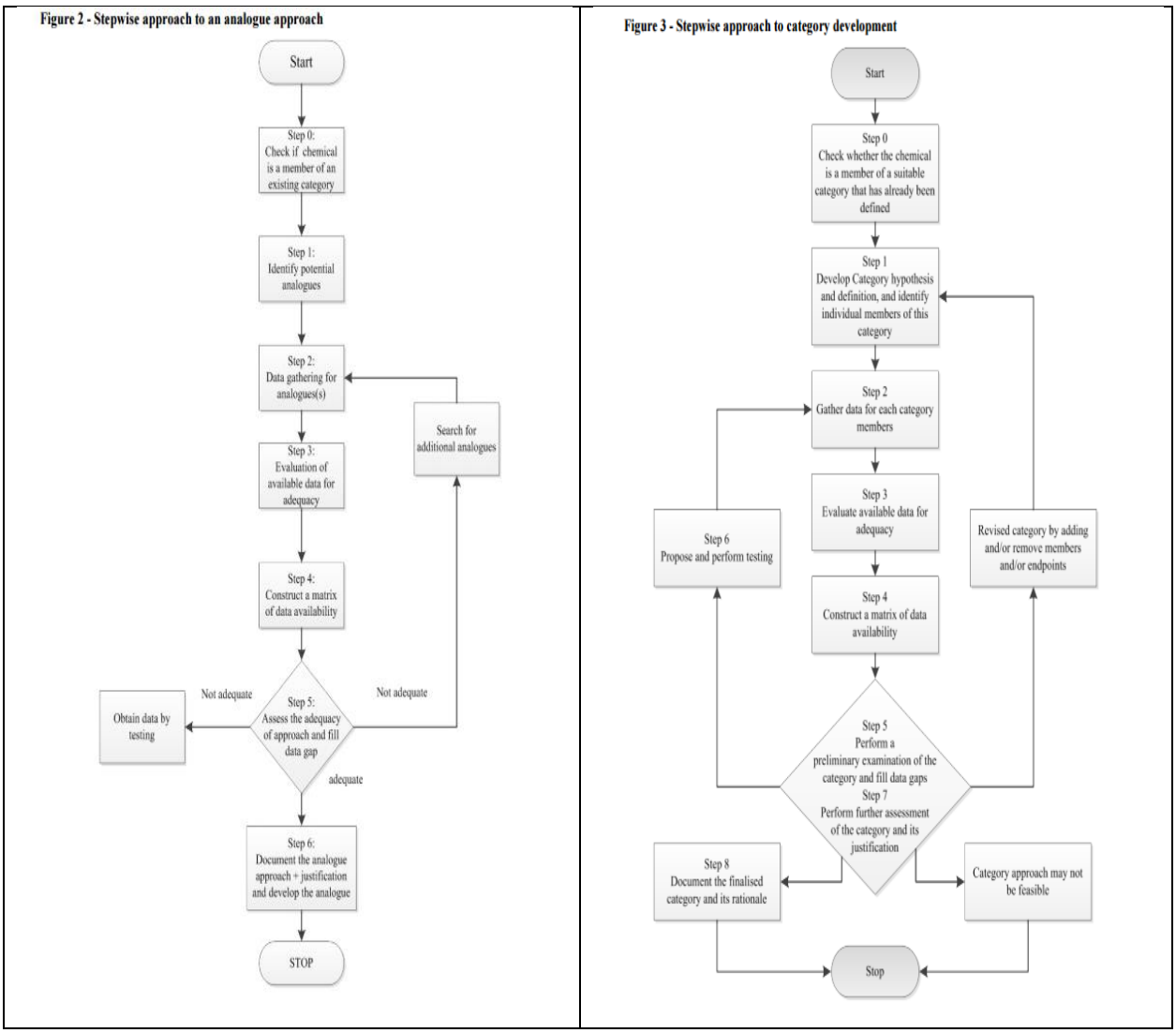

Figure 5: Overall approach by Wu et al [33]. *Reprinted from Regulatory Toxicology and Pharmacology, 56, Wu et al., A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments, pages 67-81, Copyright (2010), with permission from Elsevier*

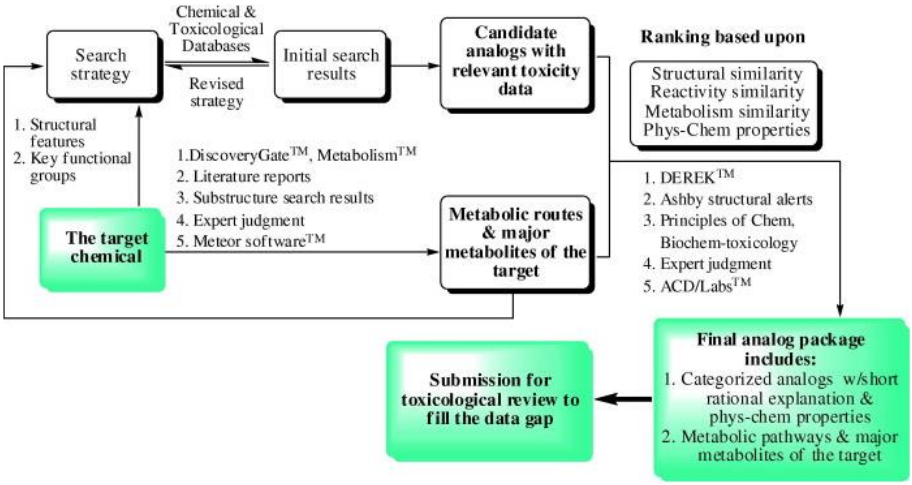

Figure 6: Analogue evaluation workflow described in Wu et al [33]. *Reprinted from Regulatory Toxicology and Pharmacology, 56, Wu et al., A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments, pages 67-81, Copyright (2010), with permission from Elsevier*

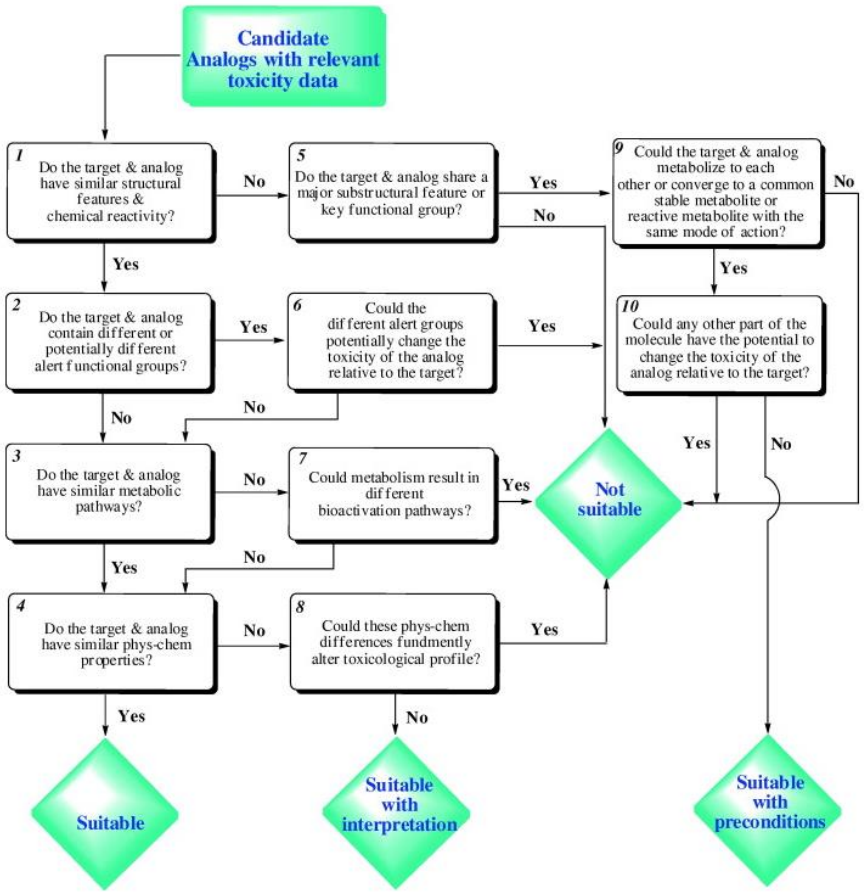

Figure 7: Workflow described in Wang et al [35]. *Reprinted from Regulatory Toxicology and Pharmacology, 63, Wang et al., Application of computational toxicological approaches in human health risk assessment. I. A tiered surrogate approach, pages 10-19, Copyright (2012), with permission from Elsevier*

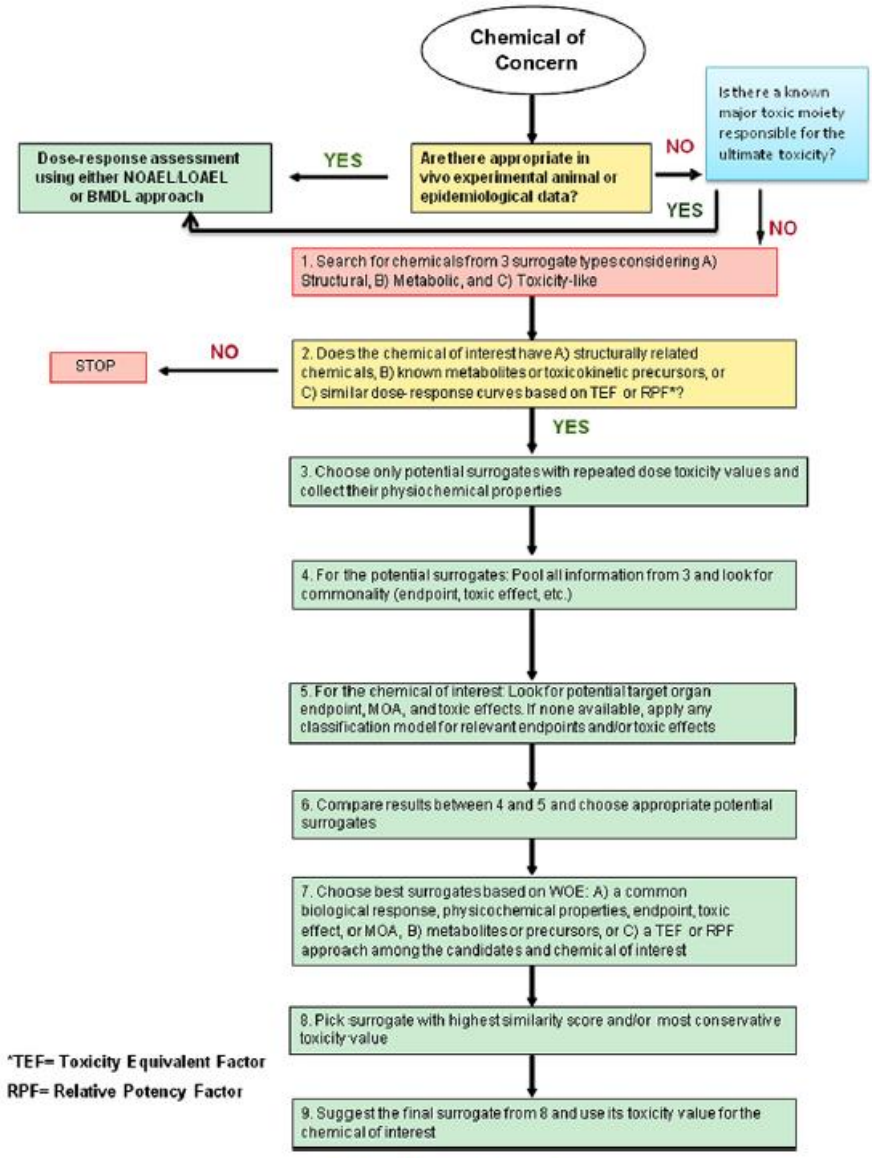

Fig. 1. Tiered surrogate approach.

Figure 8: Workflow described by Patlewicz et al [17, 20]

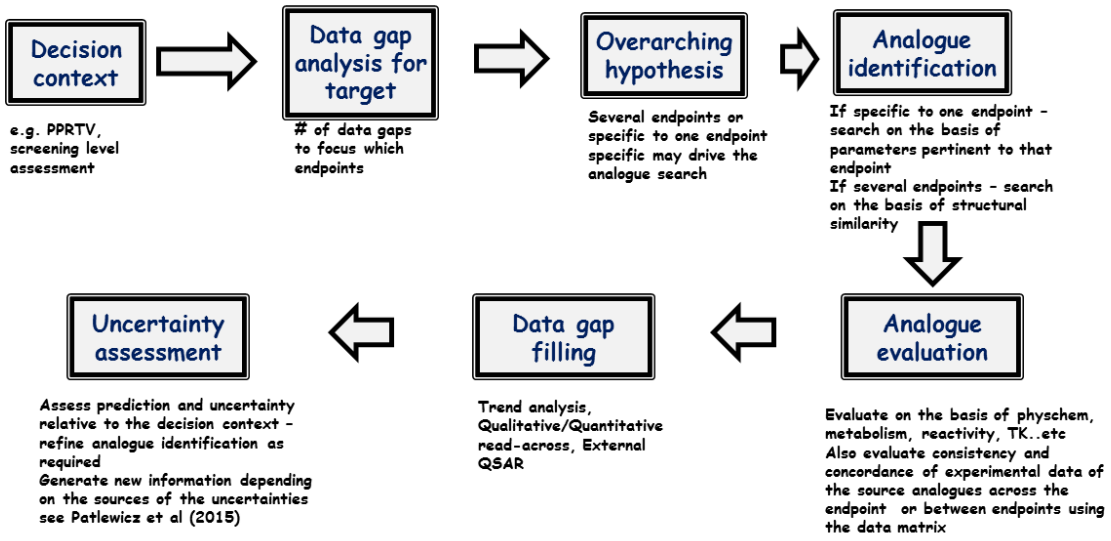

Figure 9: A Harmonised Hybrid Development and Assessment Framework