Towards grouping concepts based on new approach methodologies in chemical hazard assessment: the read-across approach of the EU-ToxRisk project

Escher, S.E.; Kamp, H.; Bennekou, S.H.; Bitsch, A.; Fisher, C.; Graepel, R.; ... ; Kroese, D.

GUEST EDITORIAL

Towards grouping concepts based on new approach methodologies in chemical hazard assessment: the read-across approach of the EU-ToxRisk project

Sylvia E. Escher[1] · Hennicke Kamp[2] · Susanne H. Bennekou[3] · Annette Bitsch[1] · Ciarán Fisher[4] · Rabea Graepel[5] · Jan G. Hengstler[6] · Matthias Herzler[7] · Derek Knight[8] · Marcel Leist[9] · Ulf Norinder[10] · Gladys Ouédraogo[11] · Manuel Pastor[12] · Sharon Stuard[13] · Andrew White[14] · Barbara Zdrazil[15] · Bob van de Water[5] · Dinant Kroese[16]

Abstract

Read-across is one of the most frequently used alternative tools for hazard assessment, in particular for complex endpoints such as repeated dose or developmental and reproductive toxicity. Read-across extrapolates the outcome of a specific toxicological in vivo endpoint from tested (source) compounds to "similar" (target) compound(s). If appropriately applied, a read-across approach can be used instead of de novo animal testing. The read-across approach starts with structural/physicochemical similarity between target and source compounds, assuming that similar structural characteristics lead to similar human hazards. In addition, similarity also has to be shown for the toxicokinetic and toxicodynamic properties of the grouped compounds. To date, many read-across cases fail to demonstrate toxicokinetic and toxicodynamic similarities. New concepts, in vitro and in silico tools are needed to better characterise these properties, collectively called new approach methodologies (NAMs). This white paper outlines a general read-across assessment concept using NAMs to support hazard characterization of the grouped compounds by generating data on their dynamic and kinetic properties. Based on the overarching read-across hypothesis, the read-across workflow suggests targeted or untargeted NAM testing also outlining how mechanistic knowledge such as adverse outcome pathways (AOPs) can be utilized. Toxicokinetic models (biokinetic and PBPK), enriched by in vitro parameters such as plasma protein binding and hepatocellular clearance, are proposed to show (dis)similarity of target and source compound toxicokinetics. Furthermore, in vitro to in vivo extrapolation is proposed to predict a human equivalent dose, as potential point of departure for risk assessment. Finally, the generated NAM data are anchored to the existing in vivo data of source compounds to predict the hazard of the target compound in a qualitative and/or quantitative manner. To build this EU-ToxRisk read-across concept, case studies have been conducted and discussed with the regulatory community. These case studies are briefly outlined.

Sylvia E. Escher and Hennicke Kamp contributed equally.

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies, e.g., of the European Chemicals Agency or of other institutions to which they are affiliated. This work reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

✉ Sylvia E. Escher
sylvia.escher@item.fraunhofer.de

✉ Bob van de Water
b.water@lacdr.leidenuniv.nl

Extended author information available on the last page of the article

Introduction

Grouping/category approaches for read-across have evolved over the last decades as important risk assessment tools to attempt filling data gaps without performing additional animal studies, e.g., starting with the OECD HPV programme (OECD 2004). Read-across is used to close data gaps, most often for complex endpoints such as toxicity after repeated exposure or developmental and reproductive toxicity (ECHA 2014).

While read-across at a first glance may appear as a straightforward and logical concept, its realization is more complicated and depends on many factors such as availability and reliability of grouped compound data. In this article, the use of new approach methodology (NAM) data

will be described to increase the confidence in a read-across approach. Moreover, we introduce the concept of biological similarity as the basis for a successful read-across approach, which goes beyond using only structural similarity.

Read-across requires a similarity assessment of the grouped compounds with regard to toxicokinetic and toxicodynamic properties. In many read-across cases, it is difficult to prove similar toxicokinetic and toxicodynamic properties within the grouped compounds, e.g., because of a sparse in vivo data matrix. It is also often a challenge to conclude on a similar adverse toxicological effect pattern, as the apical findings might vary with regard to type, severity and lowest observed adverse effect level (LOAEL) within the grouped compounds (Judson et al. 2017). The apical findings in vivo do in the majority of cases not allow for a deep insight into the mechanisms underlying the observed adverse outcomes. As an example, liver fibrosis can result from different molecular, cellular and organ responses (Cong et al. 2012; Horvat et al. 2017; Nikota et al. 2017). Here, we will illustrate how NAMs could strengthen a read-across assessment by evaluation of the toxicokinetic as well as the toxicodynamic behaviour of compounds in the human organism.

Terminology

In the following, we will use the term "read-across" to describe a category or an analogue approach as defined in the Read-Across Assessment Framework (RAAF) (ECHA 2017).

Compound(s) with relevant in vivo data will be named source compound(s) (SCs), whereas compounds, lacking experimental data are named target compounds (TCs). Within a read-across approach, endpoint data of source compounds are used to estimate the same endpoint for the target compound in a qualitative and/or quantitative way.

Depending on the similarity of the grouped compounds, the reading-across of endpoint data can be described as inter- or extrapolation. The definition of the relative similarity of grouped compounds to each other is not in the focus of this publication, therefore we will use the general term "prediction" instead of inter/extrapolation.

The term "category approach" refers to a grouping in which the data of many source compounds are used to predict the hazard of one target compound (many to one read-across) or many target compounds (many to many read-across). The term "analogue approach" refers to the prediction from one or very few source chemicals to one or many target compounds.

The properties of the grouped compounds within a category have to be similar or follow a consistent trend. As outlined in this article, read-across has to demonstrate that SC and TCs likely will cause a similar toxicological response in the human organism. The assessment of similarity, therefore,

is crucial for relevant properties usually starting with chemical similarity (structural and physicochemical parameters), but also including similarity of toxicodynamic and kinetic properties; 'similarity' here is a qualitative statement on the relative sameness of the properties under consideration.

A critical effect in this article is a primary adverse effect (as opposed to secondary effects occurring as a consequence of primary effects, e.g., extramedullary hematopoiesis, e.g., in spleen or liver and all the related haematology effects secondary to the primary effect aplastic anaemia). The definition of adversity is beyond the scope of this article. The term "lead effect" in this article refers to a critical effect which is likely to determine the point of departure (PoD) for risk assessment (e.g., NOAEL/LOAEL) in the in vivo study.

Read-across workflow

The read-across idea is simple and initially relies on the hypothesis that a (quantitative) structure activity relationship ((Q)SAR) exists. Essentially, it is assumed that structurally similar compounds will act via the same mode of action and through this cause a similar hazard in vivo. If this hypothesis is true, the hazard of a target compound can be predicted from the existing toxicity data of one or many source compounds.

In reality, the selection of source compounds and the definition of "similarity" are complex. Beside structural properties, more aspects have to be carefully considered to assess similarity with regard to toxicokinetic and toxicodynamic properties within the grouped compounds. Toxicokinetics consider absorption, distribution, metabolism and excretion (ADME) properties. Differences in the ADME properties of compounds may result in variable bioavailability, as well as variable systemic (plasma) and target organ exposure. Differences in phase 1 and phase 2 metabolism might lead to detoxification or generation of toxic/reactive metabolites.

The toxicodynamic properties of a compound result in the disturbance of an organism on different levels (e.g., biochemical pathways, signalling, tissue or organ homeostasis, interaction and binding with cellular structures) and may lead to an apical effect which can be differentiated from background variation and ultimately adversely affects the health of the organism. Toxicodynamic similarity is always connected to the endpoint under investigation (e.g., acute toxicity, specific organ toxicity, mutagenicity). It is, for example, not possible to use acute toxicity study results from source compounds to predict systemic toxicity after repeated exposure. Related endpoint data can be used as supporting information, e.g., studies with subacute exposure can be considered to pinpoint main target organs when addressing toxicity after subchronic exposure.

A read-across analysis will typically start with structural similarity and then consider data on (i) ADME and/

or physicochemical (PC) properties, (ii) the critical adverse effects observed in the in vivo studies, and (iii) the corresponding LOAEL or BMD values. In addition, (Q)SAR profilers are applied to alert for potential problematic properties/dissimilarities such as binding to (plasma)proteins, chemical reactivity, genotoxicity, etc. The evaluation of all these data leads in an iterative way to a read-across hypothesis, to a selection of the most appropriate source compounds and finally to a threshold value for the target compound.

In the majority of cases, there is no information about the mechanism(s) underlying the observed adverse effect(s) in vivo. It is, therefore, often a challenge to conclude on a similar toxicological hazard of the grouped compounds mainly based on apical findings. In vivo data inherit a certain variability, because of, e.g., small differences in the study design of the animal studies [e.g., species, strains, dose selection, dose spacing, route (Judson et al. 2017; Escher et al. 2019)] or inter-individual variability of the tested species. A better understanding of the mechanism(s) that causes an adverse outcome will, therefore, be helpful to conclude on similarity and by this, strengthen the read-across hypothesis.

Adverse outcome pathways (AOPs) might help to first illustrate and then guide the testing of underlying mechanisms, e.g., using different NAMs. The concept of AOPs conceptualises a toxicological mechanism to a series of chemical-agnostic sequential events starting with a molecular initiation event (MIE), followed by a limited number of key events and key event relationships, which lead to cellular as well as organ responses and the final adverse outcome in the organism (Leist et al. 2017). An AOP is a useful tool to structure critical steps within a complex biological process (Ankley et al. 2010). The key events are essential for the progression towards the adverse outcome and can ideally be assessed by relevant in vitro or in silico models (Villeneuve et al. 2014; Ball et al. 2016). The integration of a shared AOP, e.g., verified by NAM data can therefore strengthen a grouping approach.

Objective of EU-ToxRisk

The EU-ToxRisk project is dedicated to the development of integrated approaches to testing and assessment (IATAs), which will be used for human safety assessment of chemicals. IATAs are tailored to specific problem formulations within risk assessment. The starting point is to gather relevant existing information/data and then, where needed, additional information is generated.

In an IATA context, EU-ToxRisk evaluates different types of NAMs, e.g., human in vitro assays of different complexity ranging from high-throughput assays, 2 or 3D cellular models, human tissue slices to organ-on-a-chip approaches. The in vitro models focus on human-derived cell or tissue material to overcome species differences and are centered around the

target organs liver, kidney, lung and the neuronal target systems to predict systemic repeated dose toxicity (RDT) as well as specific in vitro models to predict developmental and reproductive toxicity (DART). EU-ToxRisk NAMs also include in silico approaches such as (Q)SAR models or physiologically based toxicokinetic (PBTK) modelling and simulation.

The overall aim of EU-ToxRisk is to replace and reduce animal testing with regard to the endpoints repeated dose and reproductive toxicity. Since this is a broad topic, we developed practical examples to gain experience on the applicability and limitations of NAMs regarding sensitivity, specificity and remaining uncertainty with respect to the addressed endpoint.

In this publication, we describe the EU-ToxRisk read-across approach, which integrates mechanistic knowledge in human hazard assessment (Leist et al. 2017). We will illustrate how MIEs and KEs from in vitro assays together with in silico model and simulation tools can be used to prove (dis)similarity or a consistent trend within a read-across assessment and by this reduce the uncertainty of the prediction for the target compound. The generated NAM data are compared to the anchoring in vivo data of the source compounds and used to extrapolate the hazard of the source compound(s) to the target compound in a qualitative and/or quantitative manner. This paper will first give a very brief overview on the current read-across guidances and will point to typical challenges in a read-across assessment using illustrative examples. We will then outline how the read-across concepts could be improved using NAMs as, e.g., being developed in the EU-ToxRisk project.

EU-ToxRisk initiated collaborations with regulators from national and European regulatory authorities such as BfR and RIVM, and ECHA, EMA, EFSA, respectively. This cooperation resulted in the improved mutual understanding of the requirements and pitfalls of read-across approaches supported by NAMs, from a scientific, academic and regulatory perspective. In our opinion, a dialogue on how best to integrate NAMs into risk assessment is one of the most important steps towards a better read-across practice and will contribute to regulatory acceptance of NAM supported read-across. In this context, an inventory of the existing guidance for NAMs-based read-across reporting was made. This resulted in a document with recommendations on the reporting templates, which are used for case studies in this project. The experiences with the reporting of case studies are foreseen for a future publication.

Overview on guidance documents and read-across workflows

A need for guidance was recognised when registering chemicals under REACH, as the majority of read-across justifications in the dossiers did not obtain acceptance upon
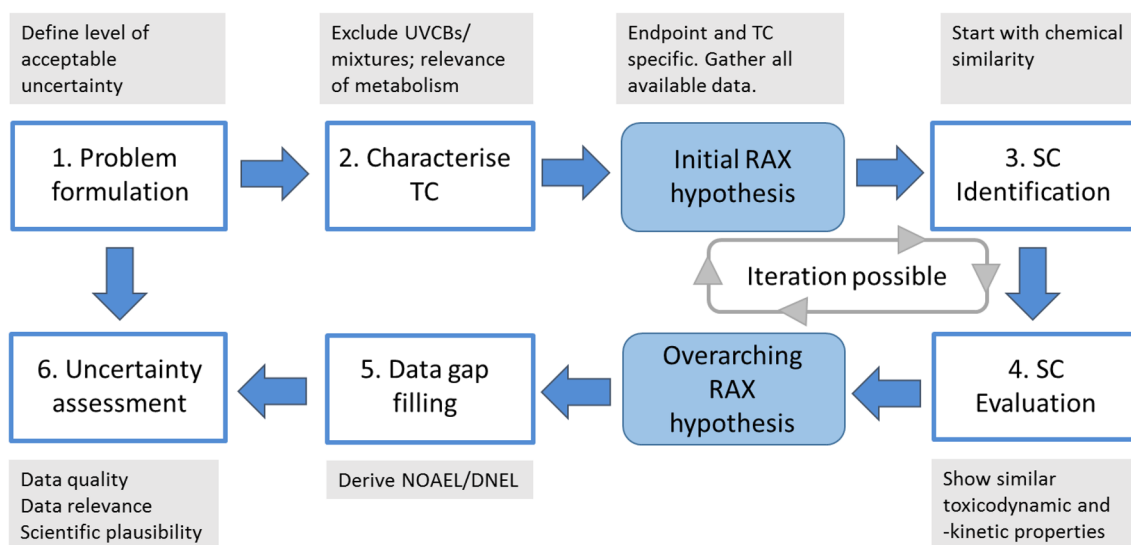
Fig. 1 Overview on the six main assessment steps of a typical read-across workflow (blue boxes), with indicated typical assessment elements (grey boxes). The problem formulation (step 1) gives the context of the read-across approach, also defining the level of acceptable uncertainty (step 5). The assessment of characteristic properties of the target compound (TC, step 2) leads to an "initial read-across hypothesis" (light blue boxes). This hypothesis guides the identifica-tion of relevant source compounds (SCs, step 3). Within the evaluation of the toxicokinetic and dynamic properties of the SC (step 4), dissimilarities might result in a refinement of the SCs, adaptation of the inclusion criteria and potentially also changing the read-across hypothesis. This iterative process results in an overarching read-across hypothesis based on which the data gap filling and uncertainty assessment is done (color figure online)

regulatory scrutiny (Ball et al. 2016). The four main reasons for rejection in disseminated compliance check decisions published by ECHA were (i) unclear substance identity of the target compound (mainly UVCBs); (ii) lack of data for analogues; (iii) read-across to inappropriate data and (iv) lack of scientific plausibility. Lack of scientific plausibility in this assessment means that data presented were not supportive of the outlined arguments, disagreed with the read-across hypothesis, contained too much uncertainty or lacked sufficient evidence/information.

One of the first guidance documents on building chemical categories and applying read-across was published in 2004 by the OECD (OECD manual for the assessment of chemicals, Sect. 3.2, 2004) and updated in 2014 (OECD 2014). In 2012, the European Center for Ecotoxicology and Toxicology of Chemicals (ECETOC) reviewed published literature and regulatory guidance documents describing the development of chemical categories (ECETOC 2012) followed by several other peer-reviewed publications on structured workflows for read-across assessments (e.g., Patlewicz et al. 2018; Schultz et al. 2015; Lizarragad et al. 2015; Blackburn and Stuard 2014). Most recently, ECHA published the read-across assessment framework (RAAF), originally developed to guide the regulators, to also support applicants with the assessment of grouped compounds (ECHA 2017).

All these approaches have in common, that they consider structural similarity as the starting point of the grouping approach. Grouped compounds have to show similar PC and

(eco)toxicological properties or these properties have to follow a consistent trend across the group, e.g., a toxicological property increases with increasing carbon side chain length.

The read-across evaluation comprises six main assessment steps, which are in some circumstances iteratively linked (Fig. 1, blue boxes). NAM can be integrated into the approach in many ways as illustrated with examples in section "A read-across workflow integrating NAMs" (Fig. 2, green boxes). The traditional six main assessment steps are:

Step 1. Problem formulation

The problem formulation accounts for the regulatory context (pharmaceuticals, chemicals, cosmetics, etc.) and use scenario, including exposure considerations, as well as for possible specific information requirements of that process/scenario, e.g., such as those provided in the Annexes VI–XI of the REACH regulation. A problem formulation is defined as "a technically oriented process that assists assessors in operationally structuring the assessment" (NAS 2009; Borghardt et al. 2015). Another aspect of problem formulation is identifying the context of the decision making based on the read-across assessment. The scope and decision context determine the amount of uncertainty that is tolerable in the final read-across result. Read-across estimates might be used, e.g.:
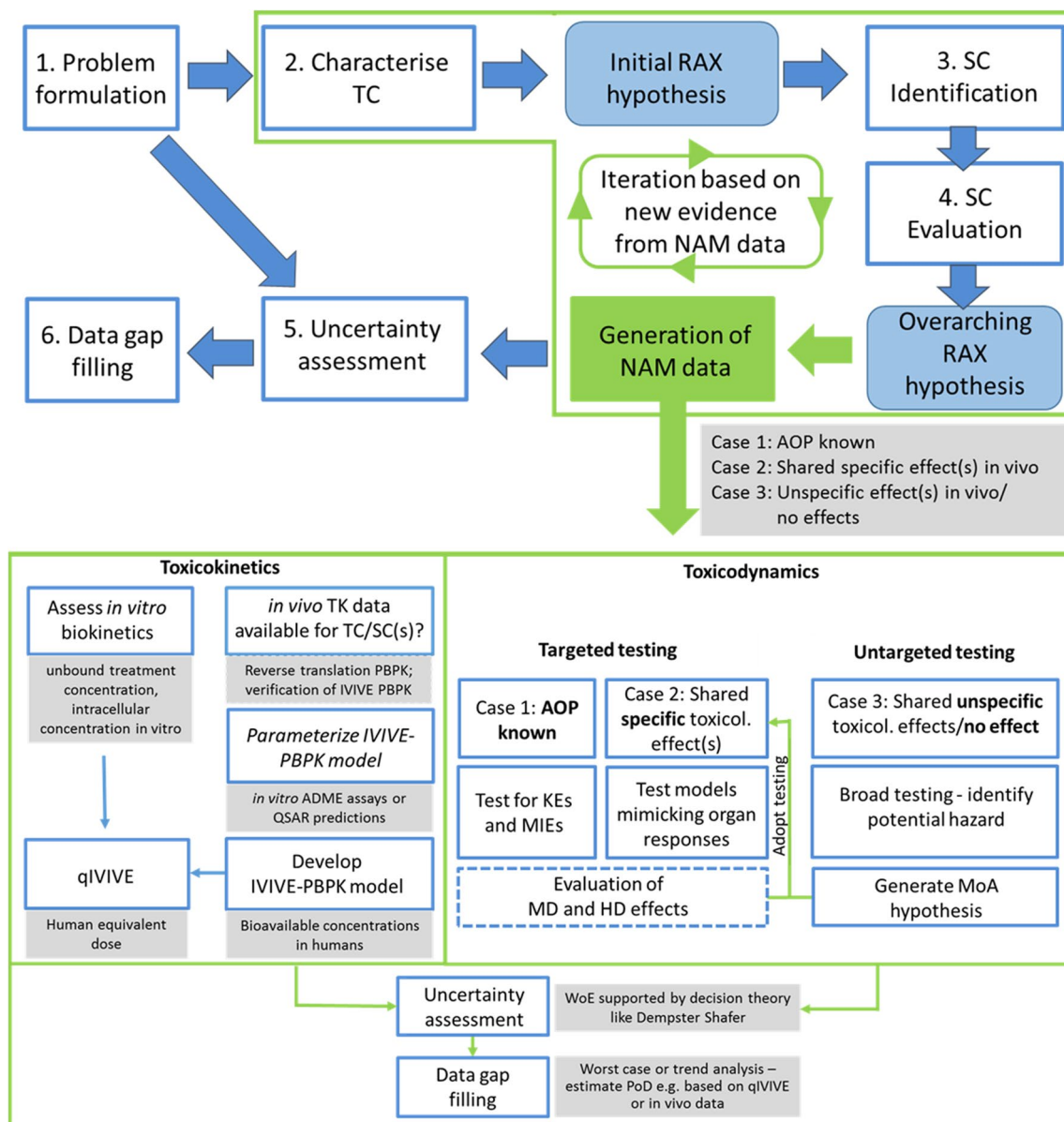
Fig. 2 Integration of NAM data into the read-across workflow. Existing NAM data can be considered within steps 2, 3 and 4 and contribute to the toxicological effect pattern of the grouped compounds leading to an overarching read-across hypothesis. For this read-across hypothesis, we distinguish three different scenarios, which determines the scope of NAM testing: Case (1): an AOP/AOP network is known—this leads to targeted testing of MIEs and selected KEs from this AOP; Case (2): analogues share (a) specific toxicological effect(s)—this leads to targeted testing of NAMs, which mimics the organ response(s); Case (3): analogues share (an) unspecific toxi- cological effect(s) or no effects up to the highest tested dose—this leads to a broad, termed untargeted, testing of NAMs with the aim to build a read-across hypothesis. The assessment of toxicokinetics aims to address (dis)similarities within the grouped compounds and to derive a human equivalent dose using a quantitative in vitro to in vivo extrapolation (QIVIVE) model. QIVIVE needs an assessment of in vitro biokinetics (left side) and an estimate of the bioavailability per compound in the human organism (right side). The latter is based on an IVIVE-PBPK model

- to establish a final threshold value for allowable lifetime exposures in a human health risk assessment;
- to establish a benchmark value for comparison to estimated exposures in a screening level assessment;
- for prioritization and screening of compounds, and/or
- for classification and labelling.

Step 2. Characterisation of the target compound(s) and development of an initial read-across hypothesis.

The read-across assessment continues with the characterisation of the target compound (TC), with the aim to generate

a first read-across hypothesis. This workflow assumes that the target compound has a defined and known structure and therefore cannot be applied to mixtures or UVCBs (substances with unknown or variable composition, complex reaction products or biological materials).

The characterisation of the target compound considers all relevant existing experimental or predicted data usually starting with structural and PC properties. PC properties and in vivo ADME data indicate the bioavailability of the TC or alert to possible bioaccumulation in the human organism. (Q)SAR models will alert for critical properties such as chemical reactivity (binding to proteins, skin sensitisation, genotoxicity) or bioaccumulation. If available, the data matrix will also comprise data for "related" in vivo endpoints. This is endpoint specific, e.g., related in vivo data for a subchronic in vivo study with oral exposure could comprise repeated dose studies with a shorter exposure period or other routes of exposure. The evaluation of the TC will lead to a first read-across hypothesis, which guides the selection of an initial set of SCs. If the TC undergoes biotransformation, the read-across hypothesis may be based on the metabolite(s), if critical, and the characterisation may have to be repeated with the metabolite(s) as target compound(s).

Step 3.   Selection of an initial set of source compounds

The selection of source compounds starts with an initial set of structurally similar compounds. Structural similarity can be assessed using different structural descriptors and algorithms or also by systematic variation of one to several key feature(s) (Croni et al. 2013). The selection of the most suitable approach is case and endpoint dependent. In any case, care needs to be taken to avoid selection bias, i.e., in- or exclusion of possible congeners should follow predefined, rigorous and transparent rules. For structurally similar source compounds, the same data types as for the TC are added to the data matrix. It is critical that some of the selected source compounds have in vivo data on the endpoint to be read-across. The in vivo data of the source compounds serve as a basis for the formulation of the overarching read-across hypothesis (result of step 4) and for anchoring the generated NAM data (steps 5 and 6).

Step 4.   Evaluation of source compounds leading to the formulation of an overarching read-across hypothesis

This step characterises the hazard of the grouped SCs to discover (dis)similarity and/or (in)consistency/ies with regard to their toxicodynamic and -kinetic properties. In traditional read-across, i.e., not integrating NAM data, analogues with relevant in vivo endpoints data will be considered, and their

in vivo ADME data (if available), PC properties and (Q)SAR predictions will be assessed. Existing relevant in vitro data might in addition be used to alert for a certain mode of action. Evaluation of the in vivo effect pattern and kinetic data might lead to a refined list of SCs (e.g., excluding SCs with dissimilar properties) and read-across hypothesis. Step 3 and step 4 may, thus, undergo several iterations (Fig. 1). Again, in- and exclusion criteria have to be described in detail to assure full transparency of the approach and to avoid a biased selection of SCs.

The chapter "A read across workflow integrating NAMs" will introduce a concept to generate NAM data based on the overarching read-across hypothesis. In contrast to generating new in vivo animal data, NAM testing is feasible within reasonable timeframes and resources for a broad set of structural analogues, allowing the assessment of effects of slight structural modifications within the category in a systematic way.

Step 5.   Data gap filling

Read-across extrapolates the in vivo data of the finally selected source compounds to the target compound. Based on problem formulation and regulatory context, the user will have to fulfil different requirements. Based on problem formulation and regulatory context, the user will have to fulfil different requirements. For example, in the context of adapting REACH standard information requirements, in vivo data of the source compounds need to allow for risk assessment and classification/labelling in the same way as the in vivo animal outcome of the target compound meant to be waived.

Data gap filling must be linked to the overarching read-across hypothesis. Acceptance of a read-across will only be achieved, if the read-across hypothesis tells a coherent "story", i.e., if all pieces of evidence combined in the read-across are linked to the problem formulation such that it becomes clear that both individually and collectively they are adequate, reliable and relevant to answer the regulatory question at hand.

A category read-across usually assumes that data for the target compound can be interpolated between source compounds of higher and lower potency. In a generic scheme, there are three general options for PoD derivation:

– a worst-case approach, basically meaning that the TC is judged to be as toxic as the most toxic compound in the group;
– a trend analysis, meaning that a consistent trend is observed, and a regression analysis can be used;
– a nearest neighbour approach, meaning that one SC is described as most similar to the TC, and only this SC's endpoint data will be read across to the TC.

Step 6. Uncertainty assessment

An uncertainty assessment needs to be carried out for all steps of the read-across approach. Excellent guidance documents are published on how to assess the uncertainty of in vivo data, and to account for data quality and data gaps in a weight of evidence approach (EFSA 2017, 2019). In addition to traditional uncertainty assessment of the available in vivo data, a read-across approach will have to provide an uncertainty assessment addressing (i) the selection of the final source compounds (e.g., outlining uncertainties arising from in/exclusion criteria), (ii) the number of source compounds, (iii) the toxicological effect pattern within the grouped compounds (e.g., addressing infrequent apical findings/or data gaps) and (iv) kinetic properties. The uncertainty assessment will probably also guide the choice of the data gap filling approach (step 5), e.g., indicating that the available data justify the worst-case but not the nearest neighbour approach. The uncertainty of the finally predicted value/property for the TC is the result of all these steps.

Uncertainty might be described in a semi-quantitative way, e.g., classifying the magnitude of uncertainty as low, moderate or high (Blackburn and Stuard 2014). Each classification will have to provide an appropriate explanation. EFSA recently proposed eight types of uncertainty assessments ranging from unqualified conclusion with no expression of uncertainty (type 1) to fully quantitative analyses using a two-dimensional probability distribution (type 8, EFSA 2019).

Read-across examples and challenges

Excellent reviews about the basic steps and recent approaches with regard to the read-across process are already available (Patlewicz et al. 2018). A major challenge in the read-across assessment is the confirmation that aside from structural and PC properties, grouped compounds also share biological properties, in particular that they will induce similar toxicological adverse effects, with different or comparable potency. This chapter discusses a few examples to illustrate where NAMs could contribute to the read-across assessment, subsequently addressing the homogeneity of in vivo data (Mangelsdorf et al. 2016) and some of the first attempts to support the read-across hypothesis by alternative data such as (i) metabolomic data from in vivo studies (van Ravenzwaay et al. 2016), (ii) NAMs for linear aliphatic alcohols (Schultz et al. 2017; Przybylak et al. 2017) and (iii) in vitro data on estrogenicity for alkylated phenols (OECD IATA case studies). The principle of so-called activity cliffs is described (Guha and Van Drie 2008). Finally,

new approaches for data integration and visualisation are briefly introduced.

Guide values for indoor air were proposed for glycol ethers and esters based on the evaluation of a data-rich category of 47 structurally very similar glycols (Mangelsdorf et al. 2016). The authors assessed in vivo data from repeated dose toxicity (RDT) studies in rodents with inhalation and oral exposure as well as reproductive studies. Although the category was relatively data rich with 147 RDT studies and 67 reproductive toxicity studies, it was a challenge to conclude on a shared toxicological effect pattern. The in vivo data showed some predominately shared toxic effects, but also a number of individual effects at several dose levels. This finding could be the result of differences in tested strains or species, study design (e.g., selection of doses and dose spacing), scope of examination or testing in different laboratories/years (Escher et al. 2019; Judson et al. 2017). As in vivo data do not directly indicate the underlying MoA, it is often a challenge to define categories based solely on apical in vivo findings. In vitro models could benefit the read-across assessment by illustrating a shared mode of action/AOPs across all grouped compounds. All compounds are metabolised to one critical metabolite, the alkoxy acid. Quantitative kinetic data were, however, not available. Oriented to precautionary principles, according to the goal to derive a guide value for indoor air, the lowest observed adverse effect concentration was used to derive a general guidance value for inhalation exposure for all members of the category. Here, NAM models could have provided more evidence on (dis)similar kinetic properties, allowing to derive compound-specific guidance values as PBPK modelling accounts for differences between compounds with regard to bioavailable concentrations in the plasma or target organs in humans.

Van Ravenzwaay et al. published an example on alternative in vivo data, which illustrates the value of metabolomics data for the substantiation of a read-across approach (van Ravenzwaay et al. 2016). The phenoxy-carboxylic acid herbicide 2-(4-chloro-2-methylphenoxy)propionic acid (MCPP) was selected as the target substance and 2-methyl-4-chlorophenoxyacetic acid (MCPA) and 2-(2,4-dichlorophenoxy) propionic acid) (2,4-DPA) as structurally closely related source substances. The evaluation of the plasma metabolome of rats treated for 28 days with the source substances indicated liver and kidney as the target organs. Metabolome evaluation of the target substance provided the same information. An overall similarity assessment of the metabolomic profiles indicated that 2,4-DPA was more closely related to the TC. The data of the 90-day oral rat study for 2,4-DPA were thus used to predict the sub-chronic toxicity of MCPP. The results of the evaluation of the overall metabolomics profile strength indicate that MCPP and MCPA have a similar potency with regard to effect, whereas 2,4-DPA was

slightly weaker in potency. The NOEL therefore would have been expected to be below the value of 2,4-DPA (<500 ppm) and in the range of that of MCPA (150 ppm). From a qualitative point of view, the predictions are very similar to the results of the actual 90-day study in rats performed with the target substance MCPP, which induced reduced food consumption and body weight gain; weight increased with concomitant clinical-pathology changes in liver/kidney and reduced red blood cell values. From a quantitative point of view, the predicted NOEL of 150 ppm is in the range of that of the actual study (NOEL 75 ppm).

To date, only few examples integrate NAM data from in vitro assays into read-across approaches. Schultz et al. predicted the NOAEL of a 90-day oral repeated dose study for a category of nine aliphatic n-alcohols, with a chain length ranging from C5 to C13 (Schultz et al. 2017). Very little experimental toxicokinetic data were available for the compounds in this category. 1-octanol was found to be rapidly absorbed after oral exposure. It is further known that some alcohols in this category form glucuronic acid conjugate and are excreted in the urine (Kamil et al. 1953). β-Oxidation is described to be the most common process in n-alkane metabolism. Data on the rate of metabolism were absent, so that compounds in this category still could have different kinetics. Two short length analogues, 1-pentanol and 1-hexanol, had experimental 90-day oral repeated dose toxicity data which exhibit qualitative and quantitative consistency. 1-heptanol, 1-undecanol and 1-dodecanol had supporting data from repeated dose toxicity studies for males with 54-day exposure (OECD TG 422). Typical findings included non-specific symptoms like decreased body weight and slightly increased liver weight which, in some cases, were accompanied by clinical chemical and haematological changes but generally without concurrent histopathological effects at the lowest observed effect level (LOEL). ToxCast data were available for the majority of the category compounds, however, to a different extent. The existing in vitro data from ToxCast and in silico prediction on nuclear receptor binding supported the read-across hypothesis, that the grouped compounds do not have an activity associated with a specific mode of action. This read-across case shows how to integrate data from in vitro and in silico models into the assessment in a qualitative way. It would, however, have been beneficial in this assessment to have (i) a consistent data matrix with similar NAM data for all compounds of the category, (ii) experimental data on ADME properties and (iii) data from in vitro tests designed to test the read-across hypothesis.

Another illustrative read-across example from the OECD IATA project is a case study developed by the US-EPA and Health Canada on the estrogenicity of alkylated phenols. Substances were screened for estrogenic potential by means of in silico and in vitro data. The data provided also aimed to estimate the in vivo point of departure doses. (Q)SAR predictions and in vitro high-throughput screening data from multiple assays were combined into a consensus prediction of estrogenic potential. Extrapolation of the in vitro bioactivity to an estimated human equivalent dose was performed through the application of reverse dosimetry. For the target substance that showed estrogenic potential, the calculated human equivalent dose was compared to effect levels from in vivo animal studies.[1]

One of the biggest challenges in read-across is to assure that a so-called "activity cliff" will not occur. An activity cliff describes a large difference in activity of paired compounds, which are similar with regard to their structural features (Guha and Van Drie 2008). This concept originates from the development of quantitative structure–property (QSPR) and structure–activity (QSAR) relationships. Activity cliffs are analysed within the training sets and test sets of these models to characterise their uncertainty. An activity cliff within the grouped compounds of a read-across evaluation will thus lead to an inappropriate prediction and failure of the read-across approach. Caution with respect to activity cliffs is probably one reason for which read-across evaluations are currently seldom accepted by authorities (Ball et al. 2016) and that for all analogues toxicodynamic and kinetic properties have to be proven similar.

We believe that NAM data can be used to alert for activity cliffs, as the testing of large series of analogues will enable the evaluation of structure–activity relationships more comprehensively compared to the current situation, where the number of source compounds is usually restricted to those with relevant in vivo endpoint data. Furthermore, mechanistic data like AOPs might be more suitable to identify activity cliffs compared to the analysis of toxicological effect patterns. New challenges follow the integration of NAM data though, for example, with respect to determining the scope of NAM testing or the calculation of human equivalent dosing, which will be described in the next section.

The assessment of biological data from NAM assays results in a need for integration and visualisation of complex, multivariate datasets. One example is the chemical–biological read-across (CBRA) approach, which intends to be a hazard classification and visualisation method. CBRA integrates chemical similarity and comparison of biological responses from multiple NAM assays into the assessment (Low et al. 2013). This approach was further developed into a more general approach, which predicts the toxicity of a target chemical using a similarity weighted activity of nearest neighbours and is now implemented within the EPA's

---

[1] http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2018)26&docLanguage=En.

CompTox Chemicals Dashboard (Shah et al. 2016; Helman et al. 2019).

EU-ToxRisk read-across framework

The EU-ToxRisk project investigates the use of NAMs in read-across and also more general hazard assessment approaches. It introduces the possibility to include biological similarity in a read-across assessment context, next to structural similarity. It also allows to verify this for the target using appropriate NAMs, when knowledge on the mechanism underlying the toxic effects in source chemicals is available, thereby reducing the uncertainty of the read-across hypothesis and the overall assessment.

In contrast to in vivo testing, NAM data can be generated within reasonable timeframes and costs for large sets of analogues within a category. The testing of a series of analogues will enable the illustration of trends or similarity in toxicokinetic and toxicodynamic properties.

The option to test a variety of NAM also results in new challenges, which are (i) how to define the scope of NAM testing; (ii) how to guide the selection of specific (relevant) NAMs, (iii) how to assess data from different NAM models, that may introduce conflicting results, and, finally, (iv) how to integrate NAM data with regard to a qualitative and/or quantitative read-across prediction.

A read-across workflow integrating NAMs

In the next section, a read-across workflow is again described, now focusing on the consequences of introducing NAMs. The workflow describes a generic read-across approach and can in principle be applied to any endpoint. The application and integration of NAMs with subsequent uncertainty assessment and data gap filling will be described in more detail in the following, together with illustrative examples (example 1 to 7).

NAMs can help to characterise the biological properties of SC and TC, thereby reducing the uncertainty of the different steps in the read-across workflow (Fig. 2). Existing in vitro data, although seldom available, can be considered within step 2 (characterisation of target compound), e.g., to alert to a specific mode of action, e.g., receptor (ant)agonism. NAMs will mainly contribute to step 4, the evaluation and confirmation of the overarching read-across hypothesis by evaluating toxicodynamic and -kinetic properties of all grouped compounds. The workflow introduces the concept that the scope of NAM testing depends on the problem formulation, the endpoint for which a read-across is performed, and the read-across hypothesis.
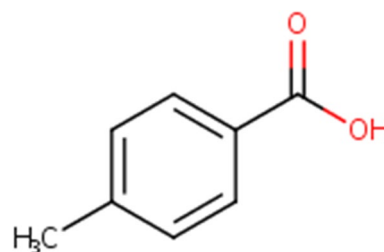

Fig. 3 2D structure of 4-methylbenzoic acid (4-MBA, CAS 99-94-5)

Step 1: problem formulation

As mentioned above, a central aspect of problem formulation is identifying the context of the decision making based on the read-across assessment. Under REACH, for example, read-across can be used to adapt the standard testing regime (Annex XI, 1.5 to the REACH regulation). Alternative models such as read-across have to provide information that is needed for classification and labelling and risk assessment.

The decision context determines the amount of uncertainty tolerable in the final read-across result and helps to select the NAM models and data, including in silico approaches, acceptable to support the decision. The read-across problem formulations can span a continuum from a restricted to a broad scope. An example of a restricted scope could be "Estimate the point of departure for a specific end-point in a repeat-dose oral exposure study for a metabolite of pesticide A". An example of a wider scope could be to "Identify and characterise the hazard of compound B".

Example 1: Problem formulation

As an example, we define 4-methylbenzoic acid (4-MBA, CAS 99-94-5) as the target compound (Fig. 3), for which we would like to predict the outcome of a subchronic oral in vivo study (this leads to step 2).

Step 2: characterization of target compound (TC) and development of an initial read-across hypothesis

The read-across assessment continues with the characterization of the target compound (TC), which is a legal requirement, e.g., under REACH. Characteristic properties of the TC will also help to generate a first read-across hypothesis (Fig. 2). To complete the picture, existing in vitro data can be considered, e.g., to alert for a specific mode of action like receptor (ant)agonism. If known, this mode of action will then have to be considered for the characterisation of the source compounds (Step 3).

Example 2: Characterization of TC leads to initial read-across hypothesis

4-MBA consists of a benzene ring, with a methyl and carboxylic acid in para position (Fig. 3). The TC is a weak acid (pK_a 4.4), water soluble (340 mg/l), not volatile (vapour pressure 6.8×10^{-5} hPa) and has a low probability to accumulate in fatty tissues ($\log P_{ow}$ 2.3). The (Q)SAR profiles of the OECD toolbox do not indicate any alert for genotoxicity or protein binding, which is used as a first (negative) indication regarding chemical reactivity.

In this example, biotransformation to a critical metabolite is assumed not to be relevant for the TC. In vivo ADME data are not available.

An initial search for in vitro data in CHEMBL shows 22 bioactivities for 4-MBA (CHEMBL ID 21708), e.g., four single protein targets like aldehyde dehydrogenase 1A1; nuclear factor erythroid 2-related factor 2; thiopurine S-methyltransferase and survival motor neuron protein. The relevance of these protein targets with regard to the toxicological effects of the TC cannot be assessed based on the available data at this stage.

In absence of any (Q)SAR alerts and relevant in vitro data, the initial read-across hypothesis for 4-MBA is structure and PC property based. We assume that aromatic carboxylic acids with similar PC properties/or following a consistent trend can be selected as a start set of analogues. Less prioritized analogues are those with additional functional groups such as halogens, thiols, nitro, nitroso, amines, aldehydes or alcohols, as well as aliphatic carboxylic acids. The risk assessor might still consider closely related functional groups like amides to be relevant. Source compounds might differ with regard to the length and/or position (ortho, meta or para) of the alkyl side chain, relative to the carboxylic acid (see Step 3). A first hypothesis on a mode of action is not possible based on the available in vivo or in vitro data.

Step 3: source compounds identification

The selection of source compounds usually starts with structural similarity. Structural similarity can be assessed using different structural descriptors and algorithms or also by systematic variation of one to several key feature(s) (Cronin et al. 2013). Three approaches can be followed:

Option 1.   *Manual selection* this method selects analogues by systematic variation of key properties of the TC.

Option 2.   *Substructure search* this method will identify all compounds that contain a certain relevant structural feature (in our example benzoic acid, Step 3 grey box). The resulting list of compounds can be structurally

very heterogeneous with regard to further substituents. This approach can be applied in cases where a substructure is known to cause a certain toxic effect. One example is anilines, which cause methemoglobinemia in vivo after biotransformation to nitrenium ions.

Option 3.   *Structure similarity* this method needs a set of descriptors, which characterise the presence/absence of structural features. The number of shared and individual structural features is then used to calculate a similarity index between the TC and each SC. A similarity threshold needs to be set to select the "most" similar analogues. It is advisable to explore different descriptors (in form of well-established fingerprints (e.g., RDKit) and algorithms (Tanimoto, Dice etc.).

Example 3: Selection of source compounds

Option 1 The systematic variation of the chain length, position and number of aliphatic side chain results in 46 potential structural analogues (Fig. 4). Benzoic acid (BA) and toluene only have one out of the two characteristic structural features of the TC. Several di- and multi-substituted analogues are possible, only di-substituted methyl and ethyl analogues are considered in this example. Similarity scores, calculated using atom pair fingerprint (RDkit) and the Tanimoto algorithm, decrease from 100% (methyl-substituted analogue in meta position) to 40% (Toluene). Toluene and 4-*tert*-butylbenzoic acid have high-quality subchronic toxicity studies with oral exposure (green), whereas 3-methyl benzoid acid and benzoic acid have supporting in vivo data from repeated dose toxicity studies with shorter study duration or inhalation exposure (Fig. 4, light green; sources of in vivo data: RepDose/ToxRef/Hess databases).

Alternatively, a similarity or substructure search can be done.

Option 2 The number of analogues depends on the inventory, in which the substructure search is performed. A search for compounds comprising the substructure "benzoic acid (BA)" in, e.g., CHEMBL reveals 110 potential analogues (Fig. 5). The list of the ten structurally most similar analogues (determined with atom pair fingerprint (RdKit) and Tanimoto algorithm) comprises alkylated benzoic acids such as 4-ethyl benzoic acid. In addition, it contains dicarboxylated (1,4-benzenedicarboxylic acid) or halogenated (5-bromo-2,3,4-trimethylbenzoic acid) analogues as well as methylester (4-methylbenzoic acid methyl ester) and one benzaldehyde (methyl 4-formylbenzoate, 4-formylbenzoic acid) (Fig. 6). 1,4-Benzenedicarboxylic acid has two in vivo studies with sub-chronic duration, none of the other analogues have in vivo data (source of in vivo data: RepDose/ToxRef/Hess databases).

single substituent		
Toluol	BA	

position and lenght of aliphatic side chain		
4-Methyl BA	3-Methyl BA 100	2-Methyl BA
4-Ethyl BA	3-Ethyl BA	2-Ethyl BA
4-Propyl BA	3-Propyl BA	2-Propyl BA
4-Isopropyl BA	3-Isopropyl BA	2-Isopropyl BA
4-Butyl BA	3-Butyl BA	2-Butyl BA
4-Isobutyl BA	3-Isobutyl BA	2-Isobutyl BA
4-tert-Butyl BA	3-tert-Butyl BA	2-tert-Butyl BA
4-Pentyl BA	3-Pentyl BA	2-Pentyl BA

di - substituted analogues		
2,4-Dimethyl BA	3,4-Dimethyl BA	2,6-Dimethyl BA
2- Ethyl-4-methyl BA	3-Ethyl-4-methyl BA	2-Ethyl-6-methyl BA
2- Methyl-4-ethyl BA	3-Methyl-4-ethyl BA	2-Methyl-6-ethyl BA
2,4-Diethyl BA	3,4-Diethyl BA	2,6-Diethyl BA
2,3-Dimethyl BA	2,5-Dimethyl BA	2,3-Diethyl BA
2-Ethyl-3-methyl BA	2-Ethyl -5-methyl BA	2,5-Diethyl BA
2-Methyl-3-ethyl BA	2-Methyl -5-ethyl BA	

Fig. 4 A overview on structurally similar compounds obtained by systematic variation of the aliphatic side chain (Example 3; Option 1). Similarity scores range from 40% for toluene to 100% for 3-meth-ylbenzoic acid (3-Methyl BA, similarity scores are indicated in orange); compounds with in vivo endpoints data are indicated in green, with supporting evidence from related in vivo endpoints in light green (color figure online)

Step 4: source compounds evaluation to derive an overarching read-across hypothesis

NAM data may provide information about shared mechanistic or kinetic properties, but are, up to now, only seldom used in read-across. In case existing in vitro data are available, the relevance and accuracy of such data with regard to the predicted in vivo endpoint need to be addressed. The evaluation of the effect pattern from all existing data (in vivo studies for the endpoint under investigation, related in vivo endpoints, in vivo ADME studies, PC properties, in silico predictions and results from relevant human in vitro models) then leads to the overarching read-across hypothesis. Within a category approach, we may have different toxicological profile situations, i.e., the category members may show (Fig. 2):

1. one common lead effect that has an established AOP (Case 1);

2. one common lead effect for which mode of action knowledge is not available (Case 2);
3. several shared lead effects, e.g., several effects in more than one target organ observed (Case 2)
4. no clear common lead effects, e.g., non-specific effects (Case 3);
5. no clear lead effect at all: an absence of effects is observed up to the highest in vivo tested dose groups, the members appear non-toxic chemicals or chemicals with very low potency and possibly non-specific in nature (Case 3).

Based on the read-across hypothesis; new NAM data can be generated (Case 1–3) to prove biological similarity of the grouped compounds.

Example 4: Evaluation of source compounds

Toxicodynamics

The five compounds with in vivo endpoint data are considered to set up the read-across hypothesis (Step 3, option 1 and 2). The analysis of PC parameters indicates that all analogues are weak acids, except for toluene. Lipophilicity increases with alkylated side chain length/absence of carboxylic acid as indicated by the logPow values. Overall, 3-methylbenzoic acid as well as benzoic acid is more similar to 4-methylbenzoic acid with regard to PC and structural properties than the other analogues. However, both compounds do not have appropriate in vivo studies for read-across, whereas 1,4-benzenedicarboxylic acid and 4-*tert*-butylbenzoic acid have subchronic in vivo studies of high quality. Toluene is not considered to be an appropriate analogue because of the missing carboxylic acid substituent (Table 1).

tert-Butylbenzoic acid shows effects in the kidney (necrosis) and testes (damage and atrophy at LOEL) followed by several mid-dose effects such as neurological symptoms (hind limb paralysis) and high-dose effects such as liver steatosis and osteoporosis (secondary effect to kidney dysfunction). 1,4-benzenedicarboxylic acid induced hyperplasia in kidney and bladder as well as a reduction in sperm counts after subchronic exposure in rats. Supporting information are available from 3-methylbenzoic acid, which showed mainly periportal hepatocellular vacuolar degeneration at LOAEL and degeneration of germ cells at high dose in male rats within a gavage study of 44 days prior to mating. A subchronic inhalation study with benzoic acid does not provide any supporting information on systemic target organs.

In addition, in vitro data might be considered to describe a toxicological effect pattern and to support the read-across hypothesis. In CHEMBL,[2] we identified 110 potential analogues with a benzoic acid substructure (Step 3; Option 2). Inspecting the pharmacological profiles of these 110 compounds (considering single protein measurements with nM unit only, Fig. 5), it becomes obvious that these in vitro data are too sparse to indicate a clear shared biological pattern. Some tendencies are visible, such as the preferred inactivity on aldehyde dehydrogenase 1A1. The ten structurally most similar analogues (Step 3, option 2) show the same trend with four out of five inactive measurements on aldehyde dehydrogenase 1A1 (Fig. 6). In this example, the observed inactivity to aldehyde dehydrogenase is considered to be of low relevance for building a read-across hypothesis. Aldehyde dehydrogenases are enzymes involved in the oxidation of aldehydes to carboxylic acids, a biotransformation that is not relevant for carboxylic acids.

Toxicokinetics

ADME properties have to be assessed to inform on differences in the bioavailability of in vivo doses. Experimental in vivo data such as plasma concentration–time profiles, Cmax or elimination half-life were not available for the five compounds. The fraction absorbed from the gut (fa) and the steady-state volume of distribution (V_{ss}; L/kg) were predicted based on PC properties and tissue composition data (Table 1). These parameters inform on the fraction of the ingested dose that will be absorbed from the gut lumen, and the extent to which the compound will distribute into tissues from the systemic circulation, respectively.

The predicted data indicate that the oral absorption of 3-methylbenzoic acid is similar to the TC, whereas greater differences are observed with respect to other analogues 4-*tert*-butylbenzoic acid and 1,4-benzenedicarboxylic acid as the result of a combination of differences in $\log P_{ow}$ and ionisation. 4-*tert*-butylbenzoic acid showed higher values for fa indicating that this compound will be more extensively absorbed from the gut. 1,4-benzenedicarboxylic acid has a lower predicted fa than the TC. Predicted V_{ss} values are comparable for all SCs and the TC; this indicates that none of these compounds will extensively distribute into tissues but will predominantly remain in the plasma instead. Beside similarity assessment, such data can be taken into consideration within the read-across assessment correcting for the LOAEL and from this to the PoD.

Overarching read-across hypothesis

The observed in vivo findings show that kidney and liver as well as reproductive cells are potential targets within the grouped compounds. The toxicological effect pattern is, however, heterogeneous and it is therefore not possible to conclude on a shared mode of action. The predicted data on ADME properties indicate that 3-methylbenzoic acid is more closely related to the TC than the other analogues. Differences in ADME properties will need to be considered when extrapolating the LOAEL of analogues to the TC. As 3-methylbenzoic acid does not have appropriate in vivo endpoint data, a one-to-one prediction is not possible.

In this example, a read-across approach based on the four identified analogues which have in vivo data would inherit a relative high uncertainty. NAM data, e.g., based on in vitro testing and in silico models could be used

to gain more insight in the MoA and the differences in ADME properties. In this example, the target organs are kidney, liver and testes, which indicate testing in NAM models that mimic the respective organ response. As the observed effects are not very specific, broader testing including, e.g., transcriptomic data is advisable (case 3, next chapter). The assessment of toxicokinetic differences would include measurement of intrinsic hepatic clearance and binding to human plasma proteins in vitro. The assessment of an effective in vitro concentration and modelling of an oral equivalent dose is explained in more detail in the chapter "toxicokinetics".

Hypothesis-driven generation of NAM data

The previous chapters describe how to define a list of source compounds and formulate a read-across hypothesis based on already existing experimental and in silico data. The next chapters outline the concept how newly generated NAM data can be used to substantiate the read-across by testing in a systematic way toxicodynamic and -kinetic properties. The scope of NAM testing is guided by the overarching read-across hypothesis (Fig. 2). Trends might be detected as NAM testing opens the floor to evaluating trends/or similarity for a broad set of structural analogues (Fig. 4, 46 analogues possible),

Table 1 Physico-chemical parameters and structural similarity scores of the source compounds with in vivo endpoint data and the target compound

| Type	Similarity score (%)	Name	Vapour pressure (hPa)	Water solubility (mg/l)	$logP_{ow}$	pK_a	fa[b]	V_{ss}^c (L/kg)
SC	100	3-Methylbenzoic acid	$3.2\,E^{-04}$	980	2.4	4.3	0.33	0.11
TC	100	4-Methylbenzoic acid	$6.8\,E^{-05}$	340	2.4	4.4	0.38	0.11
SC	90	Benzoic acid	$9.3\,E^{-04}$	3400	1.9	4.2	0.19	0.11
SC	66	4-*tert*-Butylbenzoic acid	$8.5\,E^{-04a}$	28	3.9	4.4	0.75	0.10
SC	65	1,4-Benzenedicarboxylic acid	$1.2\,E^{-05}$	15	2	3.5, 4.5	$7.2E^{-4}$	0.10
Exclude	40	Toluene	37	526	2.7	No data	–	–

Experimental values from SRC PhysProp Database

[a]Predicted value

[b]Predicted fraction absorbed from the gut following oral exposure, Simcyp mechanistic passive permeability predictor (Mech$P_{eff;}$ v18r1; Pade et al. 2017)

[c]Predicted steady-state volume of distribution (V_{ss}; method 2, Simcyp v18r1; Rodgers et al. 2005; Rodgers and Rowland 2006)

Fig. 5 A heat map showing the pharmacological effect pattern of single protein measurements for the target compound 4-methylbenzoic acid and its 110 substructural analogues extracted from ChEMBL (version 24). Red cells indicate active measurements; blue cells indicate inactive measurements (cutoff set to 10 μM). Only measurements on single protein targets and with a nM unit were considered. For visualisation purposes, targets with measurements for at least 5% of the compounds are shown (13 out of 78 targets) (color figure online)
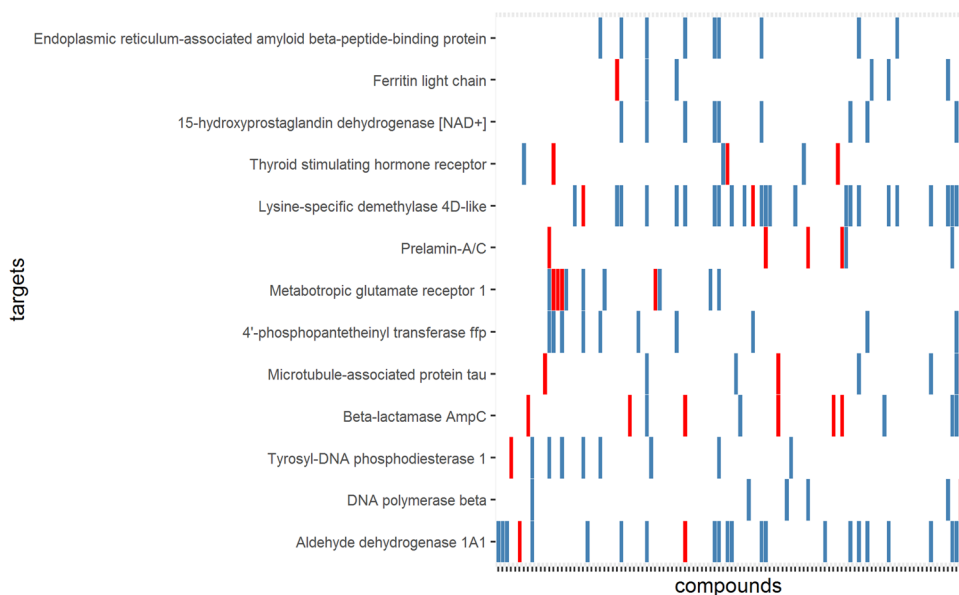
Fig. 6 A heat map showing the pharmacological pattern of 4-methylbenzoic acid and ten structurally most similar analogues extracted from ChEMBL (version 24). Red cells indicate active measurements; blue cells indicate inactive measurements (cutoff set to 10 μM) (color figure online)
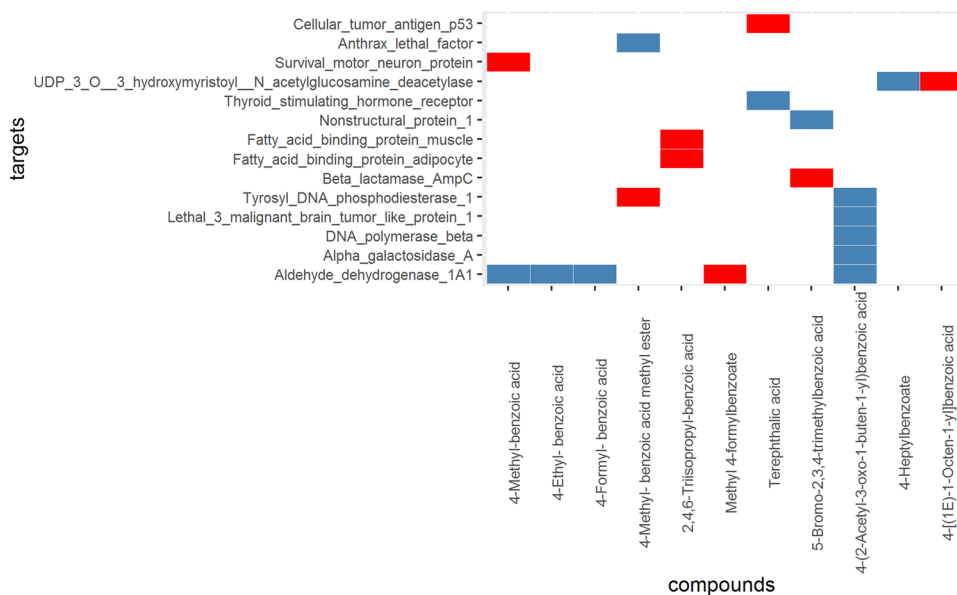
Toxicodynamics

NAM can be used to provide data on (i) test compound hazard (types of adverse outcomes expected), (ii) mode of action (pathways and targets affected), and (iii) relative potencies of effects observed in (i) and (ii). In addition, absence of a certain mechanism or effect may be tested (or low potency for a certain test endpoint be explored).

The selection of the appropriate test battery (including both experimental systems and in silico models) is a challenge that requires a detailed analysis of available data, a comprehensive definition of gaps to be filled, and a clear read-across hypothesis. From the preceding assessment steps, it is clear what kind of read-across situation we are confronted with, i.e., which source chemicals have been assessed as being adequate for this read-across substantiation, and which kind of toxicity profile is concerned. These elements define the scope of NAM testing. For the explicit definition of the test battery, it is particularly important whether a compound is expected to trigger a single specific adverse effect or rather has multiple target organs/toxicities. Specific toxicity can be defined as an (adverse) effect on a defined target structure in an animal that can be clearly defined and attributed to the tested compound (e.g., dose-dependent induction of hepatocellular necrosis). In case of a single observed specific effect, it is important whether the mode of action and the underlying AOP are known.

The selection strategy for NAMs to be used for the characterisation of toxicodynamics and kinetic properties for

read-across differs accordingly. The EU-ToxRisk framework distinguishes three cases: case 1—a shared AOP is known; case 2—shared specific apical findings are observed and case 3—no specific apical effects or no toxic effects are observed up to the highest in vivo tested dose (Fig. 2).

Case 1

If the AOP for a set of chemicals is known, NAM testing will go along this AOP and it will explore key events (KEs) or molecular initiation events (MIEs). This strategy is termed targeted testing (Fig. 2). The objective is to generate mechanism-related data for all grouped compounds, to confirm either (dis)similarity or a consistent trend. The data can then substantiate the read-across hypothesis, i.e., reduce uncertainties about potential cliffs or divergent MoAs. In vivo studies usually show several effects and the number of apical findings increases with higher dosing. It might therefore be, that more than one critical shared lead effect is observed, with, e.g., known AOP, which needs in vitro testing. Human risk assessment usually does not consider unspecific high-dose effects/or adaptive changes, e.g., weight changes or effects attributed to prominent cell death, to derive a point of departure or a classification and labelling. Those effects will have to be addressed within the evaluation of the group but will not lead to NAM testing. In cases, in which a specific adverse effect is observed in the category at doses slightly higher than the LOAEL (e.g., with a dose spacing of 2 or 3), this might also lead to additional NAM testing.
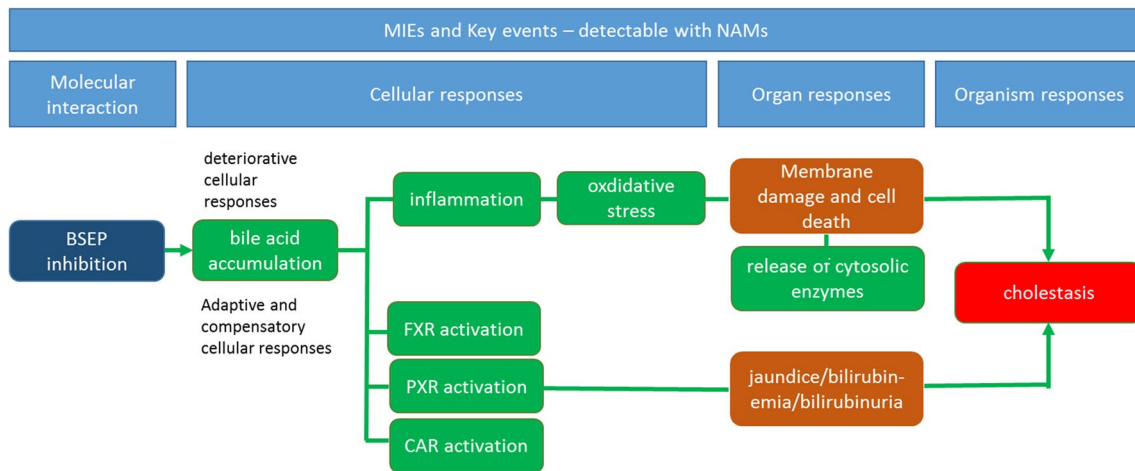
Fig. 7 AOP for liver cholestasis (adapted from Vinken et al. 2013)—the AOP is simplified to display only MIEs (dark blue) and KE (green), which can be measured by NAMs. BESP inhibition (MIE) will cause bile acid accumulation, which triggers a massive perturbation of cellular responses, accompanied by inflammation and oxidative stress. This leads to membrane damage (results are indicated in orange) and the release of cytosolic enzymes. In parallel, adaptive compensatory responses are triggered (nuclear receptor activation), which can result in jaundice/bilirubinemia and bilirubinuria at the organ level. The adverse effects in the organism is cholestasis (color figure online)

Example 5: Illustrating case 1: targeted testing of models harbouring MIEs and KEs

If the mechanism or AOP leading to the specific adverse effect is known, an in vitro test battery can be designed that tests the deregulation/activation of these specific KEs or MIEs.

One example is drug-induced liver cholestasis, for which an AOP is described (Vinken et al. 2013) (Fig. 7). A central molecular initiation event in the development of liver cholestasis is the inhibition of the bile salt export pump (BSEP). BSEP transporter protein is a prominent adenosine triphosphate-binding cassette transporter located at the canalicular pole of the hepatocyte membrane, which transports bile acids from the hepatocyte cytosol into the bile canaliculi. Inhibition of BSEP potentially causes an increase of intrahepatic bile acids with subsequent cell injury. In addition to bile acid accumulation, several KEs can be measured by NAMs on the cellular level, e.g., the induction of inflammation and oxidative stress and the activation of nuclear receptors like the pregnane X receptor (PXR), the farnesoid X receptor (FXR) and constitutive the androstane receptor (CAR, Fig. 7).

Two further examples illustrate the power of mechanism-based testing to predict adverse outcomes on the basis of NAM data: (i) the inhibition of thyroid peroxidase (TPO), the enzyme that catalyses thyroid hormone biosynthesis, leads more or less invariably to thyroid hypertrophy, and this can eventually lead to non-genotoxic tumour development (Mcclain 1992; Divi and Doerge 1996). With such clear mechanistic knowledge, TPO-based assays can predict thyroid pathology. (ii) Cardiotoxicity can be caused by the inhibition of the hERG channel, a potassium channel of high importance for the synchronisation of cardiomyocyte contraction across the whole organ. It has been shown that several drug classes, such as various neuroleptics or also modern tyrosine kinase inhibitors (TKIs) inhibit hERG and cause arrhythmias (Chaar et al. 2018). Again, the mechanistic events measurable by NAM have good predictivity for organ- or organism-level adverse outcomes.

Case 2

A specific toxicological effect may be observed such as tissue necrosis, where the underlying mechanisms/AOPs are unknown. This situation is true for the majority of adverse apical findings in animal studies.

In this case, the battery of NAM must be chosen in a way to capture all (or at least as many as possible) of the potential underlying mechanisms. A straightforward approach is to select test systems that broadly reflect target cell/organ biochemistry and physiology and to choose test endpoints that are affected by the modification of many targets and pathways (e.g., overall cell viability, or an integrated organ function such as solute transport in proximal tubule kidney cells). With the help of the EU-ToxRisk case studies (see chapter "Proof of concept—overview on ongoing case studies"), we will learn to which extent target cell/organ-specific testing is needed, having in mind that the in vitro testing battery will not aim to test all organs of the human organism for safety and risk evaluation.

Adverse effects like inflammatory responses can be induced by many different ways/AOPs. It may thus not be possible to test a MIE or very early KE, if all mechanistic information is absent. Nevertheless, there are several complex in vitro test systems which allow testing of the key processes of inflammation itself. If the source compounds induce inflammation as the primary toxicological effect (i.e., most sensitive), "targeted testing" will use models that mimic the target organ response, e.g., lung or brain slices that include inflammatory cells. Alternatively, the combination of toxicants and inflammatory mediators, such as chemokines, may be used on potential target cells to investigate response modifications by toxicants.

Case 3

If the grouped compounds share an adverse toxicological finding that is not very specific (e.g., weight loss), or if the compounds have very low toxicity, then targeted or specific testing cannot be performed. Here, NAMs would be used for broad untargeted testing to (i) either generate a read-across hypothesis based on shared in vitro effects or (ii) to prove the absence of effects (up to concentrations corresponding to those obtained in man in realistic exposure situations).

An unspecific effect would, e.g., be given by a dose-dependent significant decrease in body weight gain accompanied by a significant relative liver weight increase without any histopathological correlation and/or not clearly adverse effects like hepatocellular hypertrophy. As specific, effect-related test methods cannot be applied, models will have to be used that generate broad general data sets (such as omics data). Alternatively, broad in vitro screening batteries may be used that cover dozens to hundreds of MIEs and early KEs. These data will be used to generate a hypothesis for the underlying mechanism. Alternatively, they may be used to test whether grouped compounds show similar biological responses. It is still unclear what extent of potential biological pathways and processes needs to be covered by a screening battery, if one wants to claim the absence of an effect. EU-ToxRisk is currently starting with case studies addressing low-toxic compounds and will explore the scope of NAM testing and appropriate tiered testing strategies.

Concerning the relevance of cases 1, 2, and 3 described above, it can be expected that most read-across cases are somewhere in between case 1 and 2. The assessment of the available in vivo data will potentially show some shared toxicological findings, eventually pointing to a shared mode of action, but often some effects and/or target organs will differ. These differences prevent a conclusion on a consistently shared toxicological effect pattern within the grouped compounds. Differences in apical findings can indicate true toxicological differences, and in this case the read-across fails. They might, however, also be the result of dissimilar study designs (dose selection and spacing, tested species/strain, study duration, route) or of the variability in the in vivo data. In such cases, NAM-based data can provide clarifications. NAMs may also be used to reduce uncertainties when animal studies indicate infrequent hazardous events. This situation is difficult to interpret. Where study design is the reason underlying high variability of the in vivo models, testing appropriate in vitro or silico models for all grouped compounds (under similar conditions) may clarify the situation.

Toxicokinetics

In vivo ADME data are most often not available for industrial chemicals. Therefore, there is a need for better models describing the relationship between external dose, internal tissue or blood concentrations, or excreted amounts for both parent compounds and possible transformation products. Physiologically based pharmacokinetic (PBPK) modelling and simulation can be used to predict bioavailability and systemic/tissue exposure in humans, and model species.

EU-ToxRisk incorporates the use of in vitro to in vivo extrapolation (IVIVE) PBPK modelling. IVIVE-PBPK models are parametrised using data generated in vitro, such as intrinsic hepatic clearance ($CLint_{hep}$) in primary human hepatocytes, and plasma protein binding, to calculate the total hepatic clearance and extrapolate to the in vivo situation (Howgate et al. 2006). High-throughput assays for determining these parameters experimentally are well established, and certain parameters can be predicted using QSAR models (e.g., fraction unbound in plasma, blood to plasma ratio). While QSAR models for $CLint_{hep}$ have been published, they show only limited success, as such intrinsic hepatic clearance still represents an experimental necessity in the development of IVIVE-PBPK models. IVIVE-PBPK models in EU-ToxRisk were developed in line with the World Health Organization PBPK guidance (IPCS 2010). In addition, the approach adopted here assumes that in vivo kinetic data are available for at least one source compound to verify the predictive performance and justify model assumptions across the grouped compounds.

It is important to note that in this context, the objective of PBPK modelling and simulation is not the fully mechanistic recovery of the toxicokinetics of the read-across compounds, but to establish models for the comparison of systemic and target organ exposure across the grouped compounds based on available data. Since a focus of NAMs is to obviate the need for experimentation in animals, additional dosing studies in preclinical species to support read-across are not conducted. However, PBPK models for the prediction and cross-species

comparison of exposures can still be developed using an IVIVE approach based on legacy in vitro data using species relevant material (i.e., primary rat hepatocytes). Alternatively, a reverse translation (Rostami-Hodjegan 2018) approach may also be employed, deriving $CLint_{hep}$ from legacy toxicokinetic data in preclinical species, based on principles of pharmaco-toxicokinetics [e.g., the well-stirred liver model (Dong and Park 2018)]. If neither species-specific in vitro data, nor in vivo toxicokinetic data are available, established predictive IVIVE-PBPK models for human exposure can be used to simulate in vivo clearance in human. Predicted in vivo clearance in humans can then be allometrically scaled to the preclinical species of interest to provide a cross-species comparison if required. The species differences in specific TK mechanisms, such as enterohepatic recirculation, must be further assessed in case these mechanisms are needed to accurately describe the available in vivo data.

IVIVE-PBPK enables the integration and evaluation of ADME properties throughout the grouped compounds and high concentration differences between grouped compounds will need to be considered in the data gap filling step, e.g., by a worst-case approach or by trend analysis. The IVIVE-PBPK model can be used to eventually derive a human equivalent dose. In this approach, the free concentration in in vitro test systems is translated to a human equivalent dose based on the relevant route of exposure, biokinetic modelling of the in vitro assays (Fisher et al. 2019), and PBPK simulation (Fig. 2). Finally, the dose of the in vivo animal study that the read-across aims to waive can be predicted based on PBPK simulation in the relevant preclinical species.

Example 7: PBPK

PBPK modelling and simulation can be useful in the RAX workflow at various points. Where NOAEL/LOAEL and toxicokinetic data from in vivo studies in the same species are available, a PBPK model can be used to predict the effective concentrations in plasma and target tissues. These predicted effective concentrations can then be used to determine the range of test concentrations to be applied in vitro. Having established effective concentrations in vitro, these can be translated to in vivo equivalent external doses using reverse dosimetry on human IVIVE-PBPK models. The example below outlines a hypothetical RAX in place of a 90-day repeat-dose toxicity study in rats to assess hepatotoxicity.

How to select the concentration range for in vitro NAM testing? In vivo rat NOAEL/LOAEL studies determined a LOAEL of 500 mg/kg bw/d for hepatic steatosis for one of the SCs (SC1) in the RAX. Toxicokinetic data

---

have also been previously generated in several rat studies, providing a concentration time profile for SC1 in rat plasma. Using these available toxicokinetic profiles, a rat PBPK model is generated based on reverse translation, calculating in vivo clearance from the observed profile and then scaling this to the intrinsic hepatic clearance to parameterise the PBPK model. The predictive performance of the rat SC1 PBPK model is then verified against remaining data not used to derive model parameters. Having established a verified rat PBPK model for SC1, the oral dosing study from which a LOAEL was determined can be simulated and the maximum unbound concentrations (Cu_{max}) in plasma and liver determined. For SC1, the unbound concentration in plasma was determined to be 2.5 mM. Based on this, a concentration range of 0.125–8 mM is selected for the in vitro NAM testing of SC/TC RAX compounds. Such a model-informed approach provides an objective, data-driven strategy for and in vitro study design.

Translate in vitro NAMs to in vivo human In the next step, it is necessary to establish a human PBPK model to translate in vitro effective concentrations of grouped compounds to human in vivo oral doses. For all SCs/TCs, data on physicochemical properties [i.e., logPow, pKa, PSA ($Å^2$), HBD], solubility and volatility, are required to parametrise the PBPK models. These data are gathered from publications or databases of experimental values, or predicted using in silico tools (e.g., QSARs). Other essential model parameters such as the fraction unbound in plasma (fu), blood-to-plasma ratio (BP), and hepatic intrinsic clearance (μl/min/10^6 cells) are determined experimentally, using established in vitro methods. Here, data on the plasma concentrations following dosing in humans (at several dosing levels) were available for one of the SCs. Using these data, the predictive performance of the PBPK model for this SC is verified and used to justify the assumptions of the modelling strategy. Verification of the predictive performance of the human PBPK model for this SC confirmed the suitability of the in vitro system used to determine CLint and the applicability of the IVIVE-PBPK approach to this group of compounds.

In vitro biokinetic modelling is used to predict the intracellular concentrations corresponding to the nominal effective concentrations determined in the NAMs in vitro (Fisher et al. 2019). Based on these predicted intracellular effective concentrations, reverse dosimetry using the human PBPK models is performed to predict oral equivalent doses (OEDs; mg/kg) in humans for all SCs/TCs. Specifically, the oral dose (mg/kg bw/day) required to achieve a target organ (hepatic) C_{max} equal to the effective intracellular concentration identified in in vitro NAMs

Table 2 Structured overview of uncertainties

Level	Type of uncertainty	Examples and comments
1	Source data	Variability, completeness and heterogeneity of raw and summary data; data base reliability, etc
2	Mode of action (MoA)	How robust is the read-across hypothesis? Uncertainty about activity cliffs, and about additional/other MoA
3	Source compounds	Uncertainties about (un)biased selection of SCs. Uncertainty arising from potentially low number of adequate SCs
4	NAM data	Refers to experimental hazard and ADME tests: readiness and validation status; acceptability criteria, and reliability of data provider. For in silico models such as QSARs aspects like relevance for the in vivo end-point, prediction accuracy and applicability domain need to be considered
5	Data integration	Integration of hazard data, mechanistic endpoints and ADME data.
6	Scope	Here the uncertainty refers to the type of study results to be predicted (e.g., comprehensive 90-day study; or single organ effect in a 28-day study)
7	Gaps/inconsistencies	Major data gaps, and contradictory data need to be considered and pointed out in overall uncertainty analysis

is calculated. The human in vivo hazard can be assessed based on the in vitro hazard data, contextualised with compound-specific toxicokinetics. Since the aim of the RAX is to waive the need for the repeat-dose study in animals, PBPK can be used to simulate the results of the waived study in terms of NOAEL/LOAEL dose-level predictions. In the absence of animal clearance data, in vivo clearance predictions for all SCs/TCs from human PBPK can be allo-metrically scaled to the relevant model species. Using this approach, a rat PBPK model for the TC was constructed and used to simulate a study with repeated dosing.

Step 5: uncertainty assessment

EU-ToxRisk explores the use of NAMs in risk assessment and in particular in a read-across context. Only few NAMs have undergone full validation and incorporation into OECD test guidelines. The NAMs used in EU-ToxRisk are mainly the so-called "non-guideline methods". The quality, relevance and predictivity of such NAMs are sometimes less clearly defined than for standard animal-based testing according to OECD test guidelines. This results in different types of uncertainties, and requires a comprehensive uncertainty assessment.

As suggested by the EFSA guidance document on "uncertainty" (EFSA 2019), we use this term here in a broad sense as "referring to all types of limitations in available knowledge that affect the range and probability of possible answers to an assessment question". Available knowledge refers here to "the knowledge (evidence, data, etc.) available to assessors at the time the assessment is conducted and within the time and resources agreed for the assessment". The term 'uncertainty' is used both to refer to a source of uncertainty, and to its impact on the conclusion of an assessment. This definition is admittedly very broad, but it reflects well the situation that uncertainty for NAM-based read-across can arise at many levels and from

many sources. It is also in line with EFSA's definition. A further sharpening is expected in the future, but at present the discipline of uncertainty research is only at its beginning. Therefore, nowadays realistic uncertainty assessment has to focus mainly on a description of uncertainty sources. This means that the assessment of different uncertainties is still mainly qualitative (semi-quantitative at best), and methods for a full uncertainty quantification still need to be developed and evaluated. Within the EU-ToxRisk project, Bayesian networks have been considered for overall uncertainty extrapolation (still under evaluation), and Dempster–Shafer analysis has been employed to combine different types of information, and to produce quantitative estimates on their combined prediction accuracies.

This chapter will briefly list types of uncertainty to be considered. Some of them refer to toxicological tests in general. However, there are uncertainties that are more pronounced when using non-guideline NAMs, and there is also a group of uncertainties specific for read-across approaches.

General uncertainties comprise the limited accuracy of methods as well as the issues linked to the method's precision (= prediction accuracy). Limited accuracy is linked to the variability of data (heterogeneity of values over time, space or different members of a population, including stochastic variability (noise). Accuracy is quantified in terms of robustness/reproducibility of the method. Limited precision is linked to the fact that a method's outcome data (even if they are highly accurate) may not correlate well with effects that are (or would be) seen in humans. Precision is quantified in terms of predictivity and relevance of the method.

Non-guideline NAMs often have undergone little formal evaluation concerning robustness, relevance and predictivity. For their use in a regulatory context, readiness criteria have been elaborated (Pamies et al. 2018; Bal-Price et al. 2018a; Hartung et al. 2019) that are used in the context of EU-ToxRisk. All methods have been documented, following an extensive questionnaire. The questions cover all issues laid out by the OECD guidance document GD211 (documentation of non-guideline methods (OECD 2017), and

Table 3 Table of the outcomes for the assays on ten source compounds and one target compounds

Id	Effect	Assay												Type
		1	2	3	4	5	6	7	8	9	10	11	12	
cmpd1	Toxic	1	1	0	1	1	0	0	1	1	1	0	1	Source
cmpd2	Toxic	0	0	0	0	0	0	0	0	0	0	1	0	Source
cmpd3	Toxic	1	1	0	1	0	1	1	0	1	1	0	1	Source
cmpd4	Toxic	0	1	0	1	1	0	0	1	0	1	0	1	Source
cmpd6	Toxic	1	1	1	1	1	1	1	1	1	1	0	1	Source
cmpd7	Non-toxic	0	0	0	0	0	0	0	0	0	0	0	0	Source
cmpd8	Non-toxic	1	1	0	1	1	0	1	0	0	1	0	1	Source
cmpd9	Non-toxic	0	0	0	1	0	0	0	1	0	0	1	0	Source
cmpd10	Non-toxic	0	0	0	1	0	0	1	0	1	1	0	0	Source
cmpd11	Non-toxic	0	0	0	1	1	0	0	1	1	1	0	0	Source
cmpd5	Unknown	1	1	0	1	1	0	0	1	0	1	0	1	Target

Note that the DST software applied for this example calculation was developed in the EU-ADR project (ICT-215847)

the information is available in a transparent way in a public database (https://eu-toxrisk.douglasconnect.com/public/).

Uncertainties specific for read-across mainly arise from the need (i) to integrate several types of information to arrive at an overall conclusion, and (ii) to establish a scientific hypothesis (read-across hypothesis) that drives the overall evaluation process. Concerning (i), it is a scientific problem not yet solved, how uncertainties from largely different types of information (e.g., validity of the read-across hypothesis; suitability of the chemical similarity measures chosen; test data from NAM; predictions of metabolism) can be combined in a quantitative way. Concerning (ii) measures for the quality of a hypothesis may be adopted from other fields. However, this would not solve the issue of translating the hypothesis quality into a toxicologically relevant uncertainty measure. At present, the description of the relevant problems and uncertainties is the state of the art (e.g., Cronin et al. 2019 and also used in EU-ToxRisk). Further advances towards quantitative measures will require massive scientific efforts and financial resources.

A structured description of uncertainties and a transparent display is an objective of EU-ToxRisk. This requires various types of uncertainties to be considered at several levels of complexity (Table 2).

Some brief notes below further reflect the different levels. A more detailed discussion would be beyond the scope of this general read-across document, and EU-ToxRisk is preparing other documents on high-quality method documentation, and on an internal validation study, examining the performance of the hazard-related NAM used for the case studies.

Level 1 refers not only to NAM data, but also to the in vivo data used as anchoring and source points. Concerning level 4 (tests), uncertainties may refer to predictivity, relevance, reliability and applicability domains. With respect

to level 5 (integration), quantitative tools are emerging such as the Dempster–Shafer analysis presented below, for the integration of different hazard prediction tests. Moreover, the traditionally more qualitative integration of ADME data with hazard data is getting more and more quantitative (Punt 2018) through the use of tools allowing quantitative in vitro to in vivo extrapolations. Level 6 has two major aspects: (i) uncertainty of potency prediction (in extreme cases, either only hazard as such is predicted, or a defined human NOEL with a measure of variance for the average population or specific subgroups is derived); (ii) uncertainty of the range of endpoints to be predicted; a specific subcase is the prediction of non-toxicity, where the uncertainty of being wrong is particularly problematic. Level 7 also includes a summarizing discussion of all other levels in a balanced weight-of-evidence (WoE) approach.

Ideally, a fully quantitative system would be available to express and compare uncertainties. This could be used to drive the improvement of the read-across procedure, and to select the most appropriate NAMs for it. At present, a tool to quantify overall read-across uncertainty in a mathematical way is not available. This is in part different for read-across as compared to (Q)SAR (Cronin et al. 2019). Statistical QSAR models can be tested and validated by the use of test data which were not part of the training data, while read-across relies on expert judgement for the description, weighing and integration of very different types of uncertainties. Rule-based QSAR models, however, also include expert knowledge, nevertheless their performance can be tested with test data.

The most common approach to document uncertainty for the different levels described above is a description of the situation, followed by a WoE judgement to classify uncertainty as low, medium or high (Blackburn and Stuard 2014). However, at least for some types of uncertainty, (semi)

Table 4 Overview on assay performance: Each test is characterised by the true positive (tp)/true negative (tn) rate, the false positive (fp)/ false negative (fn) rate, as well as thereof derived values like sensitivity, specificity and balanced accuracy (BA)

Assay	tp	fn	fp	tn	Specificity	Sensitivity	BA
assay12	4	1	1	4	0.80	0.80	0.80
assay2	4	1	1	4	0.80	0.80	0.80
assay1	3	2	1	4	0.80	0.60	0.70
assay6	2	3	0	5	1.00	0.40	0.70
assay10	4	1	3	2	0.40	0.80	0.60
assay3	1	4	0	5	1.00	0.20	0.60
assay5	3	2	2	3	0.60	0.60	0.60
assay8	3	2	2	3	0.60	0.60	0.60
assay9	3	2	2	3	0.60	0.60	0.60
assay11	1	4	1	4	0.80	0.20	0.50
assay4	4	1	4	1	0.20	0.80	0.50
assay7	2	3	2	3	0.60	0.40	0.50

True and false positives and negatives, respectively: tp, fp, tn, fn

Specificity: $tn/(tn + fp)$; sensitivity: $tp/(tp - fn)$; BA $= (specificity + sensitivity)/2$

quantitative tools are available already now, and more progress is expected for the future. For instance, scores have been developed for test readiness (Bal-Price et al. 2018b) and this allows to quantify the accuracy and prediction uncertainties of tests. Moreover, the Dempster–Shafer theory (DST) (Shafer 1976; Dempster 1967; Rathman et al. 2018), a Bayesian-based decision theory approach, allows the fully quantitative combination of various types of test data, taking into account the individual test performances/uncertainties, and to derive likelihoods of test data being correct. Thus, DST-based algorithms can provide probability estimates based on the combined quality and reliability of the NAM data. This applies mainly to the combination of NAM hazard data. Incorporation of other categories of data (e.g., chemical similarity or ADME data) will necessitate modifications and extensions. The example described below is used to illustrate the application of DST to a read-across approach. DST needs positive compounds, e.g., all showing in vivo certain toxicity and one to several negative compounds, which do not show this toxicity in vivo. The information it provides on data reliability can support regulatory decisions.

Example 8: Quantifying combined uncertainties (data reliability) using the Dempster–Shafer theory (DST)

This read-across is performed using a set of twelve in vitro assays for which we assume that data are available from in vitro tests. In this example, ten source compounds have been tested, as well as one target compound. The biological properties of source compounds are known from in vivo studies: five of them showed an adverse effect (termed toxic), whereas five chemically related compound did not show this in vivo toxicity (termed non-toxic, Table 3). The effect pattern from the NAM data does not at a glance allow the conclusion that the target compound (cmpd 5) belongs to the toxic or non-toxic group. For instance, assays 1 + 2 would suggest that the target compound is toxic, while assays 3 + 6 suggest that it is not toxic (Table 3). This is a typical case where evidence from all assays needs to be combined in a way that includes background data as to how much we trust each given assay.

The actual study data (within this example) are used for evaluation of the test performance. This is possible here because of the high number of compounds with known toxicity and non-toxicity (in vivo). Based on this, it is possible to determine the false negatives (FN), false positives (FP), etc., for each test and to derive characteristics of the test prediction models such as the specificity, sensitivity and the balanced accuracy (BA) (Table 4). These data suggest that there should be different degrees of reliance on the data from the twelve tests, and here the DST provides an optimal tool to combine results of target compound testing (cmpd 5) in all assays, with the confidence measure on all twelve assays as obtained above (see Table 4).

The DST combines the above information but not in a classical probabilistic way (e.g., ANOVA or other hypothesis contrast tests). DST combines the given test data and assay quality estimates into a belief with respect to a 'proposition'. The proposition in this example is: "compound 5

Table 5 Results from DST on target compound

Name	Belief	Plausibility	DST prediction	Comment
cmpd5	0.915	0.926	Toxic	All assays used for prediction

is toxic". The DST calculation results in two output parameters: belief (BEL) and plausibility (PL). BEL indicates the strength of evidence in support of the proposition, on a scale of 0 (no certainty) to 1 (certainty). This means that the outcome of BEL = 0.915 (Table 5) means that there is 91.5% certainty that compound 5 is toxic.

The counter-proposition would be: "compound 5 is non-toxic". As mentioned above, some of the tests also delivered arguments to support this. The strength of belief into the counter-proposition is given by the plausibility parameter in the following way: if one subtracts PL from 1, then the resultant number is the belief that the compound is non-toxic. Within our example (see Table 5: results from DST on target compound), there is a probability of 1–0.926 = 0.074, i.e., there is 7.4% evidence that the target compound is non-toxic. If one adds up 7.4% and 91.5%, then 98.9% of outcome beliefs are covered. The remaining 1.1% is the difference in PL-BEL. In general, the term PL-BEL (here 0.011 = 1.1%) expresses the potential that the proposition is correct, beyond the certainty given by BEL. In this example, there is 91.5% certainty that the target compound is toxic. Altogether, there is a 92.6% potential that compound 5 is toxic, while there is only 7.4% counter-factual evidence (the uncertainty is 7.4–8.5%).

This example illustrates the outcome of the DST analysis. Moreover, it demonstrates how the input data are used to derive quantitative data on belief and uncertainty. It also shows the potential for more extensive use: for instance, the analysis may be performed for subsets of tests, and this could yield data on which tests contribute to certainty or uncertainty. Moreover, sensitivity analysis may be performed with this tool to identify areas that particularly contribute to uncertainty and need optimisation in the future.

Step 6: data gap filling supported by NAM data

Read-across extrapolates the data of the source compounds to the target compound. NAM data will strengthen the grouping approach by illustrating trends or similarity between the grouped compounds. NAM will indicate how far TC and SCs share common toxicological mechanisms/AOP or induce similar responses in test systems mimicking critical organ responses. PBPK modelling, being informed by suitable in vitro parameters, will help to detect differences in ADME properties and can be used to refine the selection of most relevant source compounds.

Based on the problem formulation and the regulatory context, the user will have to fulfil different requirements. For example, in a REACH context, a read-across has to provide information equivalent to that available from the waived standard in vivo assays, which essentially means that the in vivo data of the source compounds, together with the NAM data of source and target compounds have to predict the in vivo animal outcome of the target compound, all required as a basis for classification/labelling as well as risk assessment. Under REACH, the registrant will in general use the available in vivo data for the derivation of the NOAEL/PoD for the TC. In this case, NAM data of source and target compounds can be used to reduce the uncertainty of the read-across by illustrating a shared mode of action (AOPs), similar ADME properties, or a consistent trend.

Other regulatory contexts will allow for the direct replacement of in vivo data by appropriate and reliable in vitro data. As outlined above, in vitro assays can be used to derive benchmark dose levels which indicate the onset of a certain toxicological/biological effect, e.g., activation of MIEs or KEs, inflammatory processes, etc. PBPK modelling converts this effective in vitro concentration into human equivalent oral doses. The human equivalent doses provide information on a PoD for the target and source compounds and can be used as replacement for the PoDs derived from in vivo animal data.

Next challenge: biological read-across

The NAM-based process as described above would also be applicable to target and source compounds, which share the same AOP/mode of action but are structurally diverse. Such a read-across hypothesis is termed biological read-across. As described under cases 1 and 2, hazard characterisation of the known AOP/mode of action by selected relevant NAM models is feasible, as well as an educated guess on differences on internal dose levels using IVIVE-PBPK modelling.

As compared to the classical read-across based on structural similarity, here SCs are identified on the basis of similarity between SCs and TC regarding a certain biological profile, e.g., on biological activity or gene expression profiles (Guha and Bender 2012; Zhu et al. 2014).

The biological read-across concept is, however, not yet at a stage to be considered for risk assessment. It remains, for example, questionable to which extent structurally diverse compounds may have additional and dissimilar toxicological properties and how the most critical effects can be detected using NAMs. Besides targeted testing, additional NAMs will have to cover a broad enough toxicological space to ensure that critical adverse effects will not be overlooked (e.g., through omics methodologies using an appropriately representative selection of different cell systems).

Also, compared to a classical read-across based on structural similarity, it has to be noted that the uncertainty for the

purely biologically based read-across might be considered higher, thereby ignoring that biological activity is not always proportionally correlated with structure.

Proof of concept: overview on ongoing case studies

To enhance the transition of moving from assessments based on in vivo data to application of NAMs, the OECD is running the Integrated Approach to Testing and Assessment project (http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm#newcasestudies), where case studies are being developed. These case study assessments vary as they start from problem formulation under different regulations, which therefore comprise, for example, defined approaches, prioritization and hazard characterization and a handful of read-across cases.

In addition, the project EU-ToxRisk developed several case studies illustrating the applicability of NAMs within a read-across context, and is now further advancing to case studies in which analogues with anchoring in vivo data are not available, termed ab initio approaches. The majority of the case studies comprise structurally similar compounds and show how NAMs can be used to substantiate a read-across hypothesis.

The case studies always contain some analogues with in vivo endpoint data, so that predictivity and accuracy of the NAM data can be verified. For the same reason, some structurally relatively similar compounds are included, that do not show the shared toxicological effect pattern/AOP in the in vivo data which determines the read-across hypothesis. NAM data will be used to better define the boundaries of the categories, also showing absence or decrease of toxicity within the grouped compounds. In addition, the use of NAMs for biological read-across is under investigation. The EU-ToxRisk case studies are briefly summarized in the following section.

Microvesicular liver steatosis: a read-across case study with branched carboxylic acids

19 (un)branched aliphatic carboxylic acid is tested in selected in vitro assay systems for their ability to induce MIEs or KEs, which are described in an AOP network for liver steatosis. IVIVE-PBPK modelling is used to calculate in vivo equivalent oral doses using the most sensitive in vitro outcome per compounds. The Dempster–Shafer decision theory was used to quantify the uncertainty associated with the combination of a variety of in vitro results and furthermore helped to identify the minimal amount of assays needed for the overall conclusion.

Read-across-based filling of developmental and reproductive toxicity data gap for methyl hexanoic acid (MHA)

MHA has a data gap for developmental and reproductive toxicity. We used five structurally related two-branched aliphatic carboxylic acids that have this data to inform on MHA in a category approach. We also included less structurally related carboxylic acids as positive and negative controls, and tested all for (neuro)developmental toxicity in a battery consisting of zebrafish embryo test, mouse embryonic stem cell test, iPSC-based neurodevelopmental model, and a series of CALUX Reporter assays, that we combined with toxicokinetic models to calculate effective cellular concentrations and associated in vivo exposure doses, to identify MHA's toxicity gap profile. This in vitro and in vivo data were analysed using various statistical approaches, to conclude on the developmental and reproductive toxicity for MHA, as well as to quantify the uncertainty in the data.

Liver toxicity of hydroquinones

Six hydroquinones or resorcinols are tested for their ability to induce oxidative stress via redoxcycling in several in vitro systems. This oxidative stress is considered to be the mode of action leading to adverse liver effects in anchoring in vivo studies. The main experimental challenge in this case study turned out to be the instability of phenol derivatives in in vitro assays, together with volatility of the case compounds.

Prediction of parkinsonian-like liabilities based on AOP aligned testing linked to mitochondrial toxicity

A panel of 22 pesticides that target the mitochondrial respiratory chain and inhibit complex I, II or III were evaluated for their ability and potency to induce parkinsonian-like health effects related to inhibition of complex I. In this context, the AOP that describes this adverse outcome and has been validated by the OECD (Terron et al. 2018; Bal-Price et al. 2018b) was used as a template to establish an integrated testing strategy and integrate different test methods that allow quantitative assessment of the different key events of this AOP and translation to an in vivo situation using IVIVE-PBPK modelling. We have assessed the application of such a testing strategy in a read-across approach using a small panel of structurally similar rotenoids that inhibit complex I as well as a panel of structural similar strobilurins that inhibit complex III.

Peroxisome proliferation and kidney toxicity of herbicides

The phenoxy acetic/propionic acid herbicides form a group of structurally similar herbicides that have been shown to

induce similar systemic toxicity in rat studies. Main toxicological effects observed are liver toxicity due to peroxisome proliferation as well as kidney toxicity associated with oxidative stress. Inhibition and/or saturation of renal tubular transport has been linked to a prolongation of compound elimination, thus extending the duration of bioavailability in the blood. Within case study 5, different test systems and read-outs (e.g., CALUX reporter gene assays, HepG2 metabolomics and stress response, RPTEC/TERT1 stress response as well as transcriptomics in the different cell systems) will be used to show biological similarity in vitro, which can be used for a NAM-based read-across. Further environmentally and clinically relevant peroxisome proliferators (such as DEHP and its active metabolite, fibrates or glitazones) have been included into the testing programme. The first experimental phases have been finalized and data from the CALUX assays, HepG2 metabolomics and stress responses show that the biological effects observed can be linked to the toxicological mode of action in the liver.

Prediction of pulmonary fibrosis: a read-across case study with diketones

Several aliphatic, short chain α, β and ψ-diketones and two ketones are tested for their ability to induce interstitial pulmonary fibrosis. In vitro models like precision cut lung slices and primary bronchial epithelial cells are exposed via air–liquid application using the Fraunhofer Expo-Cube. QIVIVE will be used to translate the in vitro effect concentration to a human equivalent dose, which can be used as starting point for risk assessment.

Parabens

The parabens case study is an example in the field of repeated dose systemic toxicity and is realized as a collaboration between EU-ToxRisk and Cosmetics Europe. This case study explores whether NAMs can be used in a read-across for low-toxicity compounds with low general toxicity and weak endocrine activity. Parabens with existing safety reviews widely used in Cosmetics with available data (legacy, internal exposure data) with a dermal route of exposure were selected. Data from methyl-, ethyl- and butyl-parabens are used in read-across to fill this data gap for reproductive toxicity for propylparaben.

Different systemic endpoints are evaluated quantitatively (based on dose response) including repeated dose general target organ toxicity, reproductive toxicity, and developmental toxicity. This assessment includes evaluation of these related systemic endpoints for the target and source chemicals and utilizes traditional in vivo data as well as data from new approach methodologies (NAM). The NAM data are used with the aim to add to the weight of evidence for a scientifically robust read-across.

Drug-induced liver injury

In this case study, a test system was established that determines the probability of hepatotoxicity associated with specific oral doses and blood concentrations of test compounds. The technique can be applied to test whether structurally similar compounds would increase the risk of hepatotoxicity to a similar extent.

References

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730–741

Ball N, Cronin MT, Shen J, Blackburn K, Booth ED, Bouhifd M, Donley E, Egnash L, Hastings C, Juberg DR, Kleensreuer N, Kroese ED, Lee AC, Luechtefeld T, Maertens A, Marty S, Naciff JM, Palmer J, Pamies D, Penman M, Richarz AN, Russo DP, Stuard SB, Patlewicz G, van Ravenzwaay B, Wu S, Zhu H, Hartung T (2016) Toward good read-across practice (GRAP) guidance. Altex 33(2):149–166

Bal-Price A, Hogberg HT, Crofton KM, Daneshian M, FitzGerald RE, Fritsche E, Heinonen T, Hougaard Bennekou S, Klima S, Piersma AH, Sachana M, Shafer TJ, Terron A, Monnet-Tschudi F, Viviani B, Waldmann T, Westerink RHS, Wilks MF, Witters H, Zurich MG, Leist M (2018a) Recommendation on test readiness criteria for new approach methods in toxicology: exemplified for developmental neurotoxicity. Altex 35(3):306–352

Bal-Price A, Leist M, Schildknecht S, Tschudi-Monnet F, Pain A, Terron (2018b) Adverse Outcome Pathway on Inhibition of the mitochondrial complex I of nigro-striatal neurons leading to parkinsonian motor deficits. In: OECD Series on Adverse Outcome Pathways 7, OECD Publishing, p 184. https://searchworks.stanford.edu/view/12844208

Blackburn K, Stuard SB (2014) A framework to facilitate consistent characterization of read across uncertainty. Regul Toxicol Pharmacol 68(3):353–362

Borghardt JM, Weber B, Staab A, Kloft C (2015) Pharmacometric models for characterizing the pharmacokinetics of orally inhaled drugs. AAPS J 17(4):853–870

Chaar M, Kamta J, Ait-Oudhia S (2018) Mechanisms, monitoring, and management of tyrosine kinase inhibitors-associated cardiovascular toxicities. Oncotargets Ther 11:6227–6237

Cong M, Iwaisako K, Jiang C, Kisseleva T (2012) Cell signals influencing hepatic fibrosis. Int J Hepatol 2012:158547

Cronin MTD, Madden J, Enoch S, Roberts D (2013) Chemical toxicity prediction: category formation and read-across. In: Issues in toxicology, royal society of chemistry. ISBN 978-1849733847

Cronin MTD, Richarz AN, Schultz TW (2019) Identification and description of the uncertainty, variability, bias and influence in quantitative structure-activity relationships (QSARs) for toxicity

prediction. Regul Toxicol Pharmacol 106:90–104. https://doi.org/10.1016/j.yrtph.2019.04.007

Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. Ann Math Stat 38(2):325–339

Divi RL, Doerge DR (1996) Inhibition of thyroid peroxidase by dietary flavonoids. Chem Res Toxicol 9(1):16–23

Dong J, Park MS (2018) Discussions on the hepatic well-stirred model: re-derivation from the dispersion model and re-analysis of the lidocaine data. Eur J Pharm Sci 124:46–60

ECETOC (2012) Category approaches Read-across, (Q)SAR. Technical Report No. 113, p 192

ECHA (2014) The use of Alternatives to Testing on Animals for the REACH Regulation. Second report under Article 117(3) of the REACH Regulation. Helsinki, Finland, European Chemicals Agency, p 131

ECHA (2017). Read-across assessment framework (RAAF). ECHA-17-R-01-EN, ISBN 978-92-9495-758-0. https://doi.org/10.2823/619212

EFSA, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Schlatter JR, Silano V, Solecki R, Turck D, Benfenati E, Chaudhry QM, Craig P, Frampton G, Greiner M, Hart A, Hogstrand C, Lambre C, Luttik R, Makowski D, Siani A, Wahlstroem H, Aguilera J, Dorne JL, Dumont AF, Hempen M, Valtueña Martínez S, Martino L, Smeraldi C, Terron A, Georgiadis N, Younes M (2017) Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J 15(8):e04971

EFSA, Hart A, Maxim L, Siegrist M, Von Goetz N, da Cruz CH, Merten C, Mosbach-Schulz O, Lahaniatis M, Smith A, Hardy A (2019) Guidance on communication of uncertainty in scientific assessments. EFSA J 17(1):5520. https://doi.org/10.2903/j.efsa.2019.5520)

Escher SE et al (2019) Impact of study parameter differences on time extrapolation factors. Regul Toxicol Pharmacol (**submitted**)

Fisher C, Simeon S, Jamei M, Gardner I, Bois YF (2019) VIVD: virtual in vitro distribution model for the mechanistic prediction of intracellular concentrations of chemicals in in vitro toxicity assays. Toxicol Vitro 58:42–50

Guha R, Bender A (eds) (2012) Computational approaches in cheminformatics and bioinformatics. Wiley-Blackwell, Hoboken. https://www.wiley.com/en-us/Computational+Approaches+in+Cheminformatics+and+Bioinformatics-p-9780470384411

Guha R, Van Drie JH (2008) Structure—activity landscape index: identifying and quantifying activity cliffs. J Chem Inf Model 48(3):646–658

Hartung T, De Vries R, Hoffmann S, Hogberg HT, Smirnova L, Tsaioun K, Whaley P, Leist M (2019) Toward good in vitro reporting standards. Altex 36(1):3–17

Helman G, Shah I, Williams AJ, Edwards J, Dunne J, Patlewicz G (2019) Generalized read-across (GenRA): a workflow implemented into the EPA CompTox Chemicals Dashboard. Altex 36(3):462–465

Horvat T, Landesmann B, Lostia A, Vinken M, Munn S, Whelan M (2017) Adverse outcome pathway development from protein alkylation to liver fibrosis. Arch Toxicol 91(4):1523–1543

Howgate EM, Rowland Yeo K, Proctor NJ, Tucker GT, Rostami-Hodjegan A (2006) Prediction of in vivo drug clearance from in vitro data. I: impact of inter-individual variability. Xenobiotica 36(6):473–497

IPCS (2010) Characterization and application of physiologically based pharmacokinetic models in risk assessment. IPCS harmonization project document, no. 9. Geneva, World Health Organization, p 92

Judson RS, Martin M, Patlewicz G, Wood CE (2017) Retrospective mining of toxicology data 1 to discover multispecies and chemical class effects: anemia as a case study. Regul Toxicol Pharmacol 86:74–92

Kamil IA, Smith JN, Williams RT (1953) Studies in detoxication. XLVI. The metabolism of aliphatic alcohols; the glucuronic acid conjugation of acyclic aliphatic alcohols. Biochem J 53(1):129–136

Leist M, Ghallab A, Graepel R, Marchan R, Hassan R, Bennekou SH, Limonciel A, Vinken M, Schildknecht S, Waldmann T, Danen E, van Ravenzwaay B, Kamp H, Gardner I, Godoy P, Bois FY, Braeuning A, Reif R, Oesch F, Drasdo D, Hohme S, Schwarz M, Hartung T, Braunbeck T, Beltman J, Vrieling H, Sanz F, Forsby A, Gadaleta D, Fisher C, Kelm J, Fluri D, Ecker G, Zdrazil B, Terron A, Jennings P, van der Burg B, Dooley S, Meijer AH, Willighagen E, Martens M, Evelo C, Mombelli E, Taboureau O, Mantovani A, Hardy B, Koch B, Escher S, van Thriel C, Cadenas C, Kroese D, van de Water D, Hengstler JG (2017) Adverse outcome pathways: opportunities, limitations and open questions. Arch Toxicol 91(11):3477–3505

Lizarragad LE, Patlewicz IS, Ball GN, Boogaard PJ, Becker RA, Hubesch B (2015) Building scientific confidence in the development and evaluation of read-across. Regul Toxicol Pharmacol 72(1):117–133

Low Y, Sedykh A, Fourches D, Golbraikh A, Whelan M, Rusyn I, Tropsha A (2013) Integrative chemical-biological read-across approach for chemical hazard classification. Chem Res Toxicol 26(8):1199–1208

Mangelsdorf I, Kleppe SN, Heinzow B, Sagunski H (2016) Indoor air guide values for glycol ethers and glycol esters—a category approach. Int J Hyg Environ Health 219(4–5):419–436

Mcclain RM (1992) Thyroid-gland neoplasia—nongenotoxic mechanisms. Toxicol Lett 64–5:397–408

NAS (National Research Council) (2009) Science and decisions: advancing risk assessment. The National Academies Press, Washington, DC. https://doi.org/10.17226/12209. ISBN 978-0-309-38814-6

Nikota J, Banville A, Goodwin LR, Wu D, Williams A, Yauk CL, Wallin H, Vogel U, Halappanavar S (2017) Stat-6 signaling pathway and not Interleukin-1 mediates multi-walled carbon nanotube-induced lung fibrosis in mice: insights from an adverse outcome pathway framework. Part Fibre Toxicol 14(1):37. https://doi.org/10.1186/s12989-017-0218-0

OECD (2004) OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure Activity Relationship Models

OECD (2014) Manual for the Assessment of Chemicals. (http://www.oecd.org/env/ehs/risk-assessment/manualfortheassessmentofchemicals.htm)

OECD (2017) Guidance document for describing non-guideline. In: Vitro test methods. https://doi.org/10.1787/9789264274730-en

OECD IATA case studies (http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm#Project)

Pade D, Jamei M, Rostami-Hodjegan A, Turner DB (2017) Application of the MechPeff model to predict passive effective intestinal permeability in the different regions of the rodent small intestine and colon. Biopharm Drug Dispos. https://doi.org/10.1002/bdd.2072

Pamies D, Bal-Price A, Chesné C, Coecke S, Dinnyes A, Eskes C, Grillari R, Gstraunthaler G, Hartung T, Jennings P, Leist M, Martin U, Passier R, Schwamborn JC, Stacey GN, Ellinger-Ziegelbauer H, Daneshian M (2018) Advanced good cell culture practice for human primary, stem cell-derived and organoid models as well as microphysiological systems. Altex 35(3):353–378. https://doi.org/10.14573/altex.1710081

Patlewicz G, Cronin MTD, Helman G, Lambert JC (2018) Navigating through the minefield of read-across frameworks: a commentary perspective. Comput Toxicol 6:39–54

Punt A (2018) Toxicokinetics in Risk Evaluations. Chem Res Toxicol 31(5):285–286

Rathman JF, Yang C, Zhou H (2018) Dempster–Shafer theory for combining in silico evidence and estimating uncertainty in chemical risk assessment. Comput Toxicol 6:16–31

Rodgers T, Rowland M (2006) Physiologically based pharmacokinetic modelling 2: predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. J Pharm Sci 95:1238–1257. https://doi.org/10.1002/jps.20502

Rodgers T, Leahy D, Rowland M (2005) Physiologically based pharmacokinetic modeling 1: predicting the tissue distribution of moderate-to-strong bases. J Pharm Sci 94:1259–1276. https://doi.org/10.1002/jps.20322

Rostami-Hodjegan A (2018) Reverse translation in PBPK and QSP: going backwards in order to go forward with confidence. Clin Pharmacol Ther 103(2):224–232

Schultz TW, Amcoff P, Berggren E, Gautier F, Klaric M, Knight DJ, Mahony C, Schwarz M, White A, Cronin MT (2015) A strategy for structuring and reporting a read-across prediction of toxicity. Regul Toxicol Pharmacol 72(3):586–601

Schultz TW, Przybylak KR, Richarz A-N, Mellor CL, Escher SE, Bradbury SP, Cronin MTD (2017) Read-across of 90-day rat oral repeated-dose toxicity: a case study for selected n-alkanols. Comput Toxicol 2:12–19

Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton. ISBN 978-0691100425

Shah I, Liu J, Judson RS, Thomas RS, Patlewicz G (2016) Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information. Regul Toxicol Pharmacol 79:12–24

Terron A, Bal-Price A, Paini A, Monnet-Tschudi F, Bennekou SH, Members EWE, Leist M, Schildknecht S (2018) An adverse outcome pathway for parkinsonian motor deficits associated with mitochondrial complex I inhibition. Arch Toxicol 92(1):41–82

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

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, Landesmann B, Lettieri T, Munn S, Nepelska M, Ottinger MA, Vergauwen L, Whelan M (2014) Adverse outcome pathway development II: best practices. Toxicol Sci 142(2):321–330

Vinken M, Landesmann B, Goumenou M, Vinken S, Shah I, Jaeschke H, Willett C, Whelan M, Rogiers V (2013) Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. Toxicol Sci 136(1):97–106

Zhu H, Zhang J, Kim MT, Boison A, Sedykh A, Moran K (2014) Big Data in Chemical Toxicity Research: the use of high-throughput screening assays to identify potential toxicants. Chem Res Toxicol 27:1643–1651

Affiliations

Sylvia E. Escher[1] · Hennicke Kamp[2] · Susanne H. Bennekou[3] · Annette Bitsch[1] · Ciarán Fisher[4] · Rabea Graepel[5] · Jan G. Hengstler[6] · Matthias Herzler[7] · Derek Knight[8] · Marcel Leist[9] · Ulf Norinder[10] · Gladys Ouédraogo[11] · Manuel Pastor[12] · Sharon Stuard[13] · Andrew White[14] · Barbara Zdrazil[15] · Bob van de Water[5] · Dinant Kroese[16]

[1] Fraunhofer Institute for Toxicology and Experimental Medicine (ITEM), Hannover, Germany

[2] BASF SE, Ludwigshafen, Germany

[3] National Food Institute, Technical University of Denmark (DTU), Copenhagen, Denmark

[4] Certare UK Ltd, Sheffield, UK

[5] Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

[6] Leibniz Research Centre IfADo, Dortmund, Germany

[7] German Federal Institute for Risk Assessment (BfR), Berlin, Germany

[8] Hazard Directorate, European Chemicals Agency, Helsinki, Finland

[9] University of Konstanz, Konstanz, Germany

[10] Stockholm University, Kista, Sweden

[11] L'Oréal Rechearch & Innovation, Aulnay-Sous-Bois, France

[12] Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Universitat Pompeu Fabra, Barcelona, Spain

[13] The Procter and Gamble Company, Mason, OH, USA

[14] Unilever PLC, Bedford, UK

[15] Divison of Drug Design and Medicinal Chemistry, University of Vienna, Vienna, Austria

[16] TNO Innovation for Life, Zeist, The Netherlands