Expert Workshop “Dealing with Uncertainty of Non-Test Methods under REACH”

23-24 September 2010, ECHA Conference centre, Annankatu 18, Helsinki

Background Document

I. Introduction

A “non-test method” refers to any non-experimental method or approach that can be used to provide data for the assessment of chemicals. Data, produced by a non-test method, are called “non-test data”. Non-test methods include QSAR models and read-across/grouping approaches and can be used to predict in a quantitative or a qualitative manner the physicochemical, biological, i.e. (eco)toxicological, and environmental fate properties of substances from knowledge of their chemical structure and other properties. Both QSARs and read-across/grouping approaches are based on the principle that the properties of substances, including their biological activities, depend on their chemical structure and hence can be predicted from it (similar substances have similar properties)

I.1 Non-test methods in regulatory frameworks

Non-test methods can conditionally be used to make adequate prediction of the properties of substances, comparable to that derived from experimental test methods. Nevertheless, the application of the non-test methods requires understanding of their potential pitfalls, so they can be used reliably to make useful predictions.

The formal assessment of applicability of QSAR models on a routine basis for regulatory purposes has for the last decade received more attention than read-across/grouping. The application of QSAR for regulatory purposes was stimulated when principles for validation of QSAR models were discussed and agreed at OECD level. QSAR models can be extremely powerful tools, albeit currently with acknowledged limitations in predicting long-term toxicological properties ‘from scratch’. Nevertheless, they are already widely used and have increasing applications and utility for predicting a number of physicochemical and (eco)toxicological

2 http://www.oecd.org/dataoecd/33/37/37849783.pdf
properties and also for supporting read-across/grouping approaches under many regulatory frameworks\(^3\), \(^4\).

The grouping of substances is a more intuitive but less formalised approach, which has a longer history of regulatory application and it has been used widely by regulatory authorities in the last 30-40 years. For example in the late 80’s, there was enough accumulated experience by the U.S E.P.A to group chemicals with shared chemical and toxicological properties into categories, facilitating reviews for both submitters and U.S E.P.A reviewers by benefiting from the accumulated data and past decisional precedents.

Currently grouping and read-across approaches are believed to provide a suitable basis for data gap filling for regulatory purposes\(^5\), providing that certain conditions for successful read-across are satisfied\(^6\). These conditions are developed further in the remaining document.

1.2 Non-test methods in the REACH context

Regulation (EC) 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) sets out as one of its aims the generation of information on chemicals to assist in identifying hazardous properties and recommendations about risk management measures to prevent adverse effects on human health and the environment.

One of the main reasons for developing and adopting the REACH Regulation was to fill the data gaps for the large number of substances already in use. For many of these substances there was inadequate information on the hazards they pose to human health and the environment. Filling these data gaps will enable industry to assess the hazards and the risks and implement any risk management measures that may be necessary to protect human health and the environment. Data gaps can be filled by standard testing or by alternative methods (such as \textit{in vitro} tests) if applicable and also by non-test methods (QSAR models, trend analysis and read-across/grouping of substances). Non-test methods allow reduction of time and cost of obtaining information for registration, and also avoid unnecessary testing on animals. Annex XI of the REACH regulation stipulates the conditions to be met for using these non-test methods for registration purposes (see Section III).

In order to improve and promote the use of non-test methods for REACH, the European Commission and ECHA, in cooperation with stakeholders, already

\(^3\) Mark Cronin, John Walker, Joanna Jaworska, Michael Comber, Christopher Watts, and Andrew Worth. 2003. Use of QSARs in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. Environ. Health Perspect., 111, 1376–1390.


developed guidance and practical guides on “QSARs and grouping of chemicals”\textsuperscript{7}, “How to report QSAR”\textsuperscript{8}, “How to report read-across and categories”\textsuperscript{9} and recently on “How to avoid unnecessary testing on animals”\textsuperscript{10}. ECHA also coordinates, together with OECD, the continuous development of the QSAR Application Toolbox\textsuperscript{11}, which is an important software tool to facilitate grouping and category building and to assist in the development of read-across justifications and their transparent documentation.

The prediction of intrinsic properties of substances is inevitably connected with a certain degree of uncertainty. It is however worth noting that an inherent degree of uncertainty is also associated with the quantification of hazard properties when experimental methods are used. Both in case of non-test and experimental methods the level of uncertainty varies with the quality of the techniques used.

This workshop will concentrate on how to deal with scientific uncertainties when non-test methods are used for predicting intrinsic properties within the context of the decision making process within a regulatory framework.

II. Main objectives of the workshop

The main objectives of this workshop are:

1) To identify current scientific challenges in the regulatory acceptance of non-test data and in assessing the associated uncertainty.

2) To analyse past and current uses of non-test methods and to provide examples of how to deal with scientific uncertainty for regulatory decision making.

3) To discuss case studies related to the use of non-test data in order to identify best practices for future applications.

4) To identify ways forward in dealing with assessing the uncertainty associated with the use of non-test methods.

III. Basic principles for using non-test methods under REACH

The REACH Regulation provides specific provisions for using non-test methods for registration purposes. One of the most important conditions is that the results should be adequate for classification and labelling and/or risk assessment. Moreover adequate and reliable documentation needs to be provided. Annex XI lists all the


\textsuperscript{11} http://www.oecd.org/document/54/0,3343,en_2649_34379_42923638_1_1_1,00.html
conditions that need to be fulfilled for accepting results of QSAR models, as well as similarity characteristics for grouping substances or using a read-across approach.

The validity of a QSAR is judged according to the five OECD Principles for validation of QSARs for regulatory purposes. They include in particular a defined endpoint, an unambiguous algorithm, a defined domain of applicability, appropriate measures of goodness-of-fit, robustness and predictivity, and a mechanistic interpretation, if possible. The OECD principles have also been considered in Annex XI (paragraph 1.3) under REACH.

The basis for grouping substances or using a read-across approach can be established using the similarity characteristics specified in Annex XI (paragraph 1.5) of REACH. The grouping could be based on chemical structure, e.g. a common functional group, and/or incremental changes in carbon chain length across the group, or on other common properties such as common precursors and/or breakdown products (metabolites or environmental degradation products), or a constant pattern in the changing potency of the properties across the category.

The similarity characteristics (which could also be called criteria or principles) might be used individually. However, read-across/grouping (and similarity) may be justified based on more than one similarity rule, for example both a chain length and metabolic pathway category. Multiple justifications usually increase the confidence in validity of the grouping. It will also help in identifying if the grouping approach applies to all group members for either environmental or toxicological endpoints, or both and if it is adequate for all routes and duration of exposure and type of effects.

IV. Sources of uncertainty when using read-across/grouping approaches and QSAR models

There is commonality in read-across/grouping approaches and QSAR models as both methodologies are based on the same principle. The concept of forming groups and then using measured data from a few similar substances (also called analogues) to estimate the properties for the untested members is a common application of QSAR. In addition, QSAR models are often valuable in read-across/grouping approaches for data gap filling. In such cases, the uncertainty from using QSAR models and read-across/grouping approaches should be considered together.

Uncertainty from the use of these two non-test methods for regulatory purposes may come from: i) the way the similarity characteristics, for read-across/grouping approaches, are addressed (regarding adequacy of information and justification) and/or ii) the way the OECD Principles related to the validity of QSAR models are fulfilled (regarding adequacy of documentation and validity of the model).

Under each subsection some discussion points are proposed but the list is not exhaustive. The purpose is to aid participants in reflecting on the different aspects of uncertainty and can be used when working with the case studies of this workshop at the break-out groups.
**IV.1. Substance identity information**

For QSAR models validation, it is essential to be able to describe the model by an unambiguous algorithm: it requires having sufficient information on the substance identity, especially chemical structure and information on purity if available. Transparency on the chemicals in the training set (set of chemicals with experimental data, which were used for the development of a QSAR model or trend) is also necessary. However, there is an assumption in QSAR that the substances used to build a model are pure, which is not always the case with read-across substances.

Substance identity information is also essential when read-across/grouping approaches are applied, both the target and the read-across substances. This is needed to make sure that the predicted effects are caused by the substances themselves and not from impurities or other constituents, which are not accounted for in the composition of the target substance.

**1. Discussion Point:** Importance of substance information in relation to uncertainty from the use of non-test methods

**IV.2. Structural and functional similarity**

The structural and functional similarity between the substances depends a lot on the type of the substance itself. Chemical similarity between substances with well defined composition (mono- and multi-constituent substances) is clearly easier to establish compared to the substances with unknown or variable composition, complex reaction products or biological materials (UVCBs).

**IV.2.1 Similarity between mono-constituent substances**

The presence or absence of (eco)toxicological properties can be directly linked to the presence or absence of certain functional groups in the substance. These properties are usually related to the chemical reactivity of the chemical moieties and these considerations are important when addressing the validity of non-test methods. Absence of adequate justification on structural and functional similarities can increase uncertainty and impact on regulatory decisions regarding the validity of the use of the non-test method for a specific endpoint.

One example of a source of uncertainty in predicting hazardous properties is when the substance contains an alkyl chain that is a mixture geometric, positional or functional isomers and/or stereo-isomers. Often one CAS number covers a mixture of several isomers, which may or may not be separated by the manufacturing process. An expert judgement is needed in such case to evaluate if the isomerism matters for the chemical similarity with regard to particular physico-chemical and/or (eco)toxicological property.

**2. Discussion Point:** To what extent do the different forms of isomerism matter for the estimation of properties? Which properties are most sensitive to isomerism? Is there a rule of thumb to identify endpoints, for which the uncertainty from possible isomerism is high?
IV.2.2 Similarity between multi-constituent substances

Multi-constituent substances present a different challenge for similarity assessment even if their detailed composition is known. The main issues related to the similarity between the multi-constituent substances include the presence of a “critical” constituent with higher toxicity or presence of effect, the quantity of such constituent, as well as the variability of the composition in the whole multi-constituent substance.

3. Discussion Point: To what extent the presence of a “critical” constituent is important? How much composition variability could be acceptable for similar multi-constituent substances?

IV.2.3 Similarity between UVCBs

UVCB substances may pose additional problems for the use of non-test methods since the structure and composition are often unknown, or subject to large variability. One possible solution could be the generation of a representative structure. However, when there is uncertainty regarding the whole substance, the choice of representative structure can be debatable. Another approach could be that a series of “representative structures” is generated and trends among them are analysed but to do this sufficient knowledge on the composition is needed in order to aggregate the results. Another possible confounding factor is if there are components with specific mechanism/mode of action. These are often seen as “outliers” in statistical models but could remain unnoticed unless trend analysis is applied in combination with the read-across.)

4. Discussion Point: How to identify outliers (substances with specific, different mechanism/mode of action and different toxicological profile) in case of grouping and read-across? Should they be excluded by the group for all endpoints, or only for the endpoints affected?

Reading-across between mono-constituent substances and UVCBs is an issue. While it seems reasonable to read-across from one UVCB to another similar UVCB (i.e. there is about the same level of uncertainty related to the substance identity, although this is case dependent), it is more difficult to read-across from a UVCB to a monoconstituent substance. Read-across the other way round is also uncertain, since a UVCB as a target substance might contain constituents with specific effects that are not present in the read-across mono-constituent substance.

5. Discussion Point: Is it possible to derive a generalised approach for reading-across between mono-constituent substance and UVCB?

IV.3. Common break-down products

A trend or similarity in toxicokinetic properties is considered important when assessing the acceptability of non-test methods.
For QSAR model validation it is necessary to define their domain of applicability (i.e. parametric, structural, metabolic and mechanistic domain). All these levels of domain increase certainty in the prediction from a model. A “domain of applicability” could also be defined for a group of substances without a formalised (quantitative) model being developed.

In the case of read-across/grouping approaches the similarity of the substance and the analogue(s) needs very often to be addressed in terms of similar toxicokinetic properties or trends in the toxicokinetic properties, since they are often linked to the mode of action for a specific endpoint and can have an impact on the predicted property. Read-across/grouping may not always be applicable for different routes and duration of exposure, and type of effects. Although the use of experimental evidence to address toxicokinetic properties would be preferred, very often hypothesis and/or non test data linking substance physicochemical properties, or known biochemical pathways, with potential kinetic aspects can be used to support this part of similarity. This lack of experimental evidence, however, on case by case basis may increase the uncertainty in the prediction.

6. Discussion Point: Importance of toxicokinetics in read/across/grouping approaches and uncertainty depending on the amount of information available.


When the data available allows, establishing a mode of action for a specific endpoint can be very useful to increase robustness of the predictivity of hazardous properties. Information or hypothesis building regarding similar mode of action or trends in (eco)toxicological properties should be assessed together with chemical reactivity similarities taking into account toxicokinetic information (experimental or predicted). In the future, advanced technologies such as -omics and modelling of ligand-receptor interactions could provide support to the establishment of MOA for a group of substances. Assessing all available information together (Weight of Evidence approach) can increase the level of certainty in the prediction and form the basis for acceptance for regulatory purposes.

For QSAR models, the mechanistic interpretation (if possible, before or after the model development) is important to understand the association between the chemical descriptors (e.g. hydrophobic, steric, electronic and other quantitative derivatives of the chemical structure) used and the endpoint modelled. Clearly, it is not always possible to provide mechanistic interpretation for a given QSAR model. Nevertheless, models developed on mechanistic basis often show advantages compared to models with mechanistic interpretation or no interpretation at all. Such models can provide additional information also within the read-across/grouping approach and increase certainty in the prediction of a hazardous property.

7. Discussion Point: How a common MOA for read-across/grouping purposes could be defined? How much and what type of information will be regarded as sufficient for regulatory acceptance of similarity based on common mode of action?

---

IV.5. Endpoints covered

Another requisite for QSAR models is to have a defined endpoint when addressing the validity of the model in order to ensure transparency in the endpoint being predicted.

For read-across/grouping approaches, the endpoint should be especially well defined when different types of mode of action are addressed, as well as when different routes and duration of exposure are present, and the type of effect is different (local versus systemic toxicity). Assessment of the overall data should be done in a Weight of Evidence approach to allow sound conclusions on which endpoint are covered by read-across/grouping.

The uncertainty in the experimental test data exists and this can derive from a number of sources such as experimental error or model error. The accuracy of the experimental test data should always be considered when the accuracy of the non-test data is assessed.

8. Discussion point: Recommendations regarding the endpoint coverage or the defined endpoint in dealing with uncertainty when using read-across/grouping approaches and QSAR models respectively.

V. The way forward

Experience gained through the establishment of the OECD Principles for QSAR models can serve as the basis of the future work towards a better understanding of what could be considered as good practice for read-across/grouping approaches. This could be possible only by discussion of the uncertainty and its sources and comparison with the achievable certainty from test data alone. Evaluation of all available information in a Weight of Evidence approach is believed to be able to increase the confidence in the use of non-test data. Therefore, a future framework for the application of read-across/grouping is currently needed to facilitate transparency in developing acceptable justifications and documenting of the remaining uncertainty.

VI.1. Short term goals as outcome of this WS

1. To summarise the outcomes of this workshop for participants ECHA Committees and stakeholders. This process should aim at forming a roadmap for considering ways and means to identify, document and evaluate uncertainty when using non-test data, mainly derived from grouping and read-across, for regulatory purposes.

2. The outcomes of this Workshop with relation to grouping and read-across will be brought to the attention of OECD in order to facilitate the consensus building between OECD Member Countries on the matter of using surrogate data, derived mainly from grouping and read-across, for regulatory purposes.

3. Stakeholders and public are regularly informed on this activity.
VI.2. Mid-to-long term goals

4. After the first deadline for registration under REACH, ECHA will have collected more information on the use of non-test data in registration dossiers. This information will be analysed and reported in the context of Art. 117 (paragraph 3) of the REACH Regulation.

5. This accumulated experience will be useful to verify whether the Guidance on “QSARs and grouping of chemicals”© could be revised to give more up-to-date examples and practical help.

6. The ECHA Secretariat is to consider organising of future workshops to track the progress and update the roadmap for non-test method implementation.

7. The use of the QSAR Toolbox should be encouraged and exemplified in terms of the emerging ideas from this Workshop and the subsequent developments.

8. REACH data can be used for further development and validation of non-test methods to create a better scientific basis for using non-test methods and dealing with uncertainty in risk assessment.