EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

19 January 2026
EMA/CHMP/4299/2026
Committee for Medicinal Products for Human Use (CHMP)

Reflection paper on the qualification of non-mutagenic impurities

Draft agreed by Non-Clinical Working Party (NcWP)	October 2024
Adopted by CHMP for release for consultation	02 December 2024
Start of public consultation	30 January 2025
End of consultation (deadline for comments)	30 April 2025
Agreed by NcWP	03 December 2025
Adopted by CHMP	19 January 2026

Keywords	Non-mutagenic impurities, pharmacology, toxicology, threshold of toxicological concern, read across, animal testing, in vitro testing, 3Rs

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

Reflection paper on the qualification of non-mutagenic impurities

Table of contents

List of abbreviations

ADME	Absorption, distribution, metabolism and excretion
AF	Assessment factor
AI	Artificial intelligence
AL	Acceptable level
ALARP	As Low As Reasonably Possible
AOP	Adverse outcome pathways
API	Active pharmaceutical ingredient (active ingredient without impurity)
ATMP	Advanced Therapy Medicinal Products
BMD	Benchmark dose
BMDL	Benchmark dose lower boundary
BMDU	Benchmark dose upper boundary
BMR	Benchmark response
bw	Bodyweight
CHMP	Committee for Medicinal Products for Human use
CNS	Central nervous system
CVMP	Committee for Veterinary Medicinal Products
CVS	Cardiovascular system
DNA	Deoxyribonucleic acid
DP	Drug product
DS	Drug substance (active ingredient with impurities)
DST	Dermal sensitisation threshold
EDT	Expanded decision tree
EFSA	European Food Safety Authority
E_I	Anticipated maximal systemic exposure to impurity
ELSIE	Extractables and leachables safety information exchange
E_M	Estimated systemic exposure to metabolite
EMA	European Medicines Agency
EU	European Union
GIT	Gastrointestinal tract
GLP	Good laboratory practice

HPC	High potency category
ICH	International Council for Harmonization of Technical Requirements of Pharmaceuticals for Human Use
iTTC	Internal threshold of toxicological concern
LOAEL	Lowest observed adverse effect level
LOEL	Lowest observed effect level
MCC	Matthew's correlation coefficient
MDD	Maximum daily dose
ML	Machine learning
MOC	Maximum observed concentration
MTC	Maximum theoretical concentration
NAM	New approach methodology
NcWP	Non-clinical working party
NMI	Non-mutagenic impurities
NOAEL	No observed adverse effect level
NOEL	No observed effect level
OECD	Organization for Economic Cooperation and Development
PBPK	Physiologically based pharmacokinetic
PDE	Permitted daily exposure
PK	Pharmacokinetic
PoD	Point of departure
QIVIVE	Quantitative in vitro-in vivo extrapolation
QMRF	QSAR Model Reporting Format
QPRF	QSAR Prediction Reporting Format
(Q)SAR	(Quantitative) structure activity relationship
QT	Qualification threshold
QWP	Quality working party
ROC-AUC	Receiver operating characteristic area under the curve
SWP	Safety working party
TK	Toxicokinetic
TTC	Threshold of toxicological concern
WoE	Weight of evidence

1. Executive summary

The ICH Q3A(R2) and Q3B(R2) guidelines provide a framework for qualifying Non-Mutagenic Impurities (NMI) in drug substances (DS) and drug products (DP) for registration applications but offer limited guidance on newly identified or elevated impurity levels above existing ICH Q3A or Q3B impurity qualification thresholds (QT), that are identified in post-approval procedures after non-clinical toxicology studies have been completed for initial Marketing Authorisation Applications (MAAs). This reflection paper recognises the need for an adequate safety evaluation and, in line with the 3Rs principles, particularly the replacement of tests on animals, suggests alternative strategies to in vivo animal studies for qualifying novel or elevated levels of impurities specified above ICH Q3A/Q3B QTs. ICH Q3A/Q3B states that impurities can be qualified when present in the DS/DP used in the non-clinical repeated dose toxicity studies at a level that does not impact the outcome of the toxicological assessment of the DS/DP. Furthermore, impurities may be qualified when these are also present as metabolites in animals or humans. This reflection paper provides further recommendations on how to assess metabolism data for qualifying an impurity as well as recommendations on a similarity assessment between the impurity and the Active Pharmaceutical Ingredient (API) to justify read across to existing toxicological studies with the DS.

The need for additional data depends on the level of concern. The level of concern for an NMI is affected in a multifactorial manner, including intake level, physico-chemical properties and degradability of the NMI, as well as route of administration and bioavailability, clinical conditions and target population for the medicinal product. If a need for additional data is identified, the primary source should be existing toxicological data that can be used to derive an Acceptable Level (AL). This can be based on existing impurity-specific data from the literature or data from an adequate surrogate molecule supported by read-across. The AL method estimates a product-specific safe level of exposure to impurities. If an AL above the DS or DP specification limit of an NMI can be established, the level of impurity can be considered toxicologically qualified. The AL method is based on the Permitted Daily Exposure (PDE) methodology described in ICH Q3C, but also considers bioavailability, read-across data and product-specific information. It involves selecting a point of departure (PoD) from toxicological studies and applying assessment factors (AFs).

In the absence of adequate data to derive an AL, New Approach Methodologies (NAM) can be employed, involving characterisation of chemical properties and computational toxicology tools. Combining evidence from multiple in silico tools or in vitro studies can fill knowledge gaps, with expert opinions complementing the risk assessment.

Only when alternative methods fail to provide sufficient information to establish the safety of an impurity at the proposed specification limit, and with appropriate scientific justification, should in vivo studies be considered. In this situation, this reflection paper provides specific recommendations for the design of new in vivo studies for qualification of NMIs.

Impurities in investigational medicinal products should be evaluated according to ICH M3(R2). When there is a need for additional safety data, the principles in this reflection paper can complement the approaches described in ICH M3(R2) for evaluating the safety of impurities during clinical development.

In summary, when impurity-specific safety information for an NMI is recommended, alternative strategies to gathering this information may be followed, including the use of existing toxicological data, read-across, threshold of toxicological concern (TTC), computational and in vitro approaches. This information can be used in an integrated risk assessment. A weight-of-evidence (WoE) approach that includes all aspects that determine the level of concern, could be sufficient to decide that the NMI can be considered safe at the specified level.

2. Introduction

The ICH Q3A(R2) and Q3B(R2) guidelines provide a framework for qualifying Non-Mutagenic Impurities (NMI) in DSs and DPs, including reporting, identification and qualification thresholds (QT) for impurities. These guidelines define qualification as "the process of acquiring and evaluating data that establishes the biological safety of an individual impurity or a given impurity profile at the level(s) specified." They also state that "the applicant should provide a rationale for establishing impurity acceptance criteria that includes safety considerations." According to ICH Q3A and Q3B, no further toxicological data is necessary to qualify specified impurity levels, as included in the DS or DP specifications, which are below the qualification thresholds. Furthermore, impurity levels can also be qualified when present in the DS/DP used in the non-clinical repeat dose toxicology studies at levels that do not impact the outcome of the toxicological assessment of DS/DP. In addition to ICH Q3A and Q3B, other impurity specific guidelines are in place such as ICH M7(R2), Q3D and Q3C (see Section 4 for expanded list), which address DNA reactive (mutagenic) impurities, elemental impurities and residual solvents, respectively. However, overall, these guidelines offer limited guidance on the process of qualification of newly identified or elevated impurity levels above existing ICH Q3A or Q3B impurity QTs that are identified after non-clinical toxicology studies for the API are complete, beyond the conduct of additional in vivo general toxicology studies. This reflection paper aims to provide recommendations to complement the current approaches in ICH guidelines, specifically to provide recommendations on alternative methods to additional in vivo toxicology testing with NMIs.

The ICH Q3A/Q3B guidelines recommend that additional safety testing should be considered when higher levels or new impurities, that cannot be qualified through existing non-clinical data, trigger the need for qualification. As a result, dedicated animal studies have often been conducted with the aim of qualifying these impurities, however, these studies are often designed using impurity-spiked API-batches instead of testing the neat impurity, which has several limitations.

In the impurity-spiked DS studies, the biological safety of the DP is established as a whole based on its specific impurity profile, rather than characterising the safety of the individual impurities. When toxicity is observed, it is typically not possible to determine whether it is attributable to the API or to one or more of the impurities present in the DP batch. Impurities are typically also present at much lower concentrations than the API, adding further complexity in attributing any observed effects specifically to the impurity. Thus, No Observed Adverse Effect Level (NOAEL) derived from such studies is most likely reflective of the API's safety profile rather than that of the impurity (Graham et al., 2021). While these studies may be considered valid in terms of design, they do not yield relevant information regarding the biological safety of the impurity. This undermines the scientific justification for conducting such studies and conflicts with Directive 2010/63/EU on the protection of animals used for scientific purposes, which discourages animal testing when it is unlikely to produce meaningful data. In fact, a survey amongst stakeholders reviewing in vivo studies on 467 impurities found no cases where toxic impurities were identified. In 98.7% of these studies, the impurity was present— either spiked or unspiked —at low levels in the DS (Slikkerveer et al., 2024), further highlighting the limited value of this approach. Another key observation when non-clinical studies on the DS is used for qualification of impurities is that allometric scaling is often not considered and the safety evaluation does thus not consider the biological differences between humans and animal species. In some instances, this leads to an in vivo study not being accepted by regulators to support the impurity safety evaluation as the impurity levels in the DS non-clinical study are too low when also applying corrections for allometric scaling to the NOAEL.

The approach of using impurity-spiked DS batches also stands in contrasts to the compound-specific approaches adopted for mutagenic impurities, elemental impurities and solvents, as well as the approach under development for extractables and leachables. These guidelines emphasise the

importance of evaluating the safety of individual impurities, rather than relying solely on the overall impurity profile of the DS or DP.

In cases where impurities have not been previously qualified in safety studies (i.e. novel impurities), or when higher levels of impurities need to be qualified (that were previously qualified at a lower level), the use of alternative, non-animal methodologies are recommended. Only in rare instances — where a residual safety concern cannot be resolved through other means — should the conduct of an animal study be considered. In such cases, a study with the neat impurity is more likely to provide relevant information on the impurity's safety profile due to the reasons outlined above.

This reflection paper addresses the qualification of NMI from a safety perspective when impurity levels included in the DS or DP specification exceed the QTs in ICH Q3A and Q3B. It does not address the acceptability of impurity levels from a quality perspective.

3. Scope

This reflection paper outlines the current thinking on recommended approaches for evaluating the safety of NMIs. The principles and methodologies discussed are intended to support the qualification of new or elevated impurity levels, which are specified at levels that exceed existing ICH Q3A/Q3B QTs, particularly when such impurities are identified after completion of non-clinical toxicology studies. These new or elevated impurity levels may result from any type of DS or DP related changes, such as changes in manufacturing processes. The reflection paper does not reflect further on the ALARP principle (As Low As Reasonably Possible) for the limits of impurities, but specification levels should generally be based on batch data and not only be guided by toxicologically qualified levels. In addition, this paper complements the approaches described in ICH M3(R2) for evaluating the safety of impurities during clinical development. Its aim is to supplement existing guidelines —specifically ICH Q3A(R2) and Q3B(R2)— by proposing alternative, non-animal strategies for the qualification of NMIs. These approaches may offer more impurity-specific information than traditional in vivo studies. It also acknowledges that the level of concern for impurities may vary depending on multiple factors, which in turn determine the extent of data needed, ranging from no additional data to impurity-specific experimental studies.

The scope of this reflection paper is limited to NMIs in chemically synthesised pharmaceuticals. It does not address impurities in Advanced Therapy Medicinal Products (ATMP), in herbal medicinal products, and in biological and biotechnologically derived pharmaceuticals. However, for combination products containing both chemically synthesised and biotechnologically derived components (e.g. antibody drug conjugates) the principles outlined in this paper may be considered where appropriate.

4. Legal basis and relevant guidelines

In the EU, the legal basis for qualification of drug impurities is primarily found in Directive 2001/83/EC, which sets out the general requirements for marketing authorisations in conjunction with guidance issued by European Medicines Agency (EMA) and the ICH guidelines (GL). These documents outline scientific principles, including thresholds for reporting, identification, and qualification of specific classes of impurities and of impurities in specific classes of drug substances, and the requirement to generate toxicological data or to provide evidence showing the impurity is safe when its level exceeds the set thresholds.

- non-mutagenic impurities (NMI) in chemically synthesised drug substances (ICH Q3A(R2)) and drug products (ICH Q3B(R2)),
- NMI qualification approaches during clinical development (ICH M3(R2))
- mutagenic impurities (ICH M7(R2)),
- solvents (ICH Q3C),
- elemental impurities (ICH Q3D),
- impurities in oligonucleotides (Guideline on the Development and Manufacture of Oligonucleotides (EMA/CHMP/CVMP/QWP/262313/2024),
- impurities in chemically synthesised peptides (Guideline on the Development and Manufacture of Synthetic Peptides (EMA/CHMP/CVMP/QWP/387541/2023),
- impurities in radiopharmaceuticals (Guideline on Radiopharmaceuticals – Revision 1 (EMEA/CHMP/QWP/306970/2007).

5. Key considerations

5.1. General outline for risk assessment of NMIs

The principles described in this reflection paper may be considered for the toxicological qualification of identified NMIs when the following conditions are met:

a) the specified level of the impurity in the DS or DP specification exceeds the QT as defined in ICH Q3A(R2) and Q3B(R2);

b) existing non-clinical studies with the DS do not adequately address the impurity at the identified/specified levels (e.g. the impurity was not detected or was present only at low concentrations).

The principles of this reflection paper may also be applicable when the NMI is not covered by other impurity-specific guidelines, as listed in Section 4.

The general strategy for risk assessment of NMIs is illustrated in the flow chart in Figure 1.

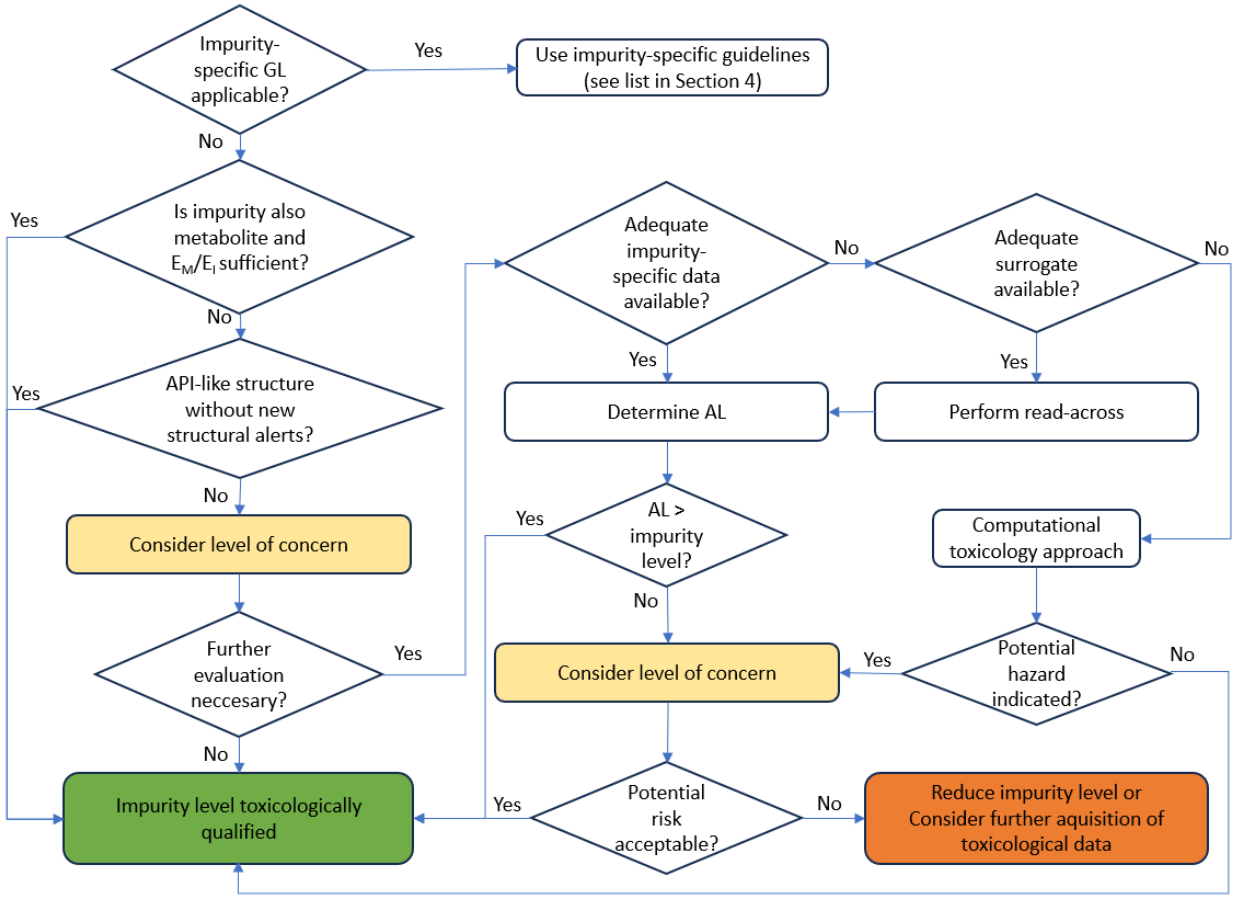
The flow chart is structured into focused areas of evaluation to support a stepwise assessment of NMIs. This approach enables the use of a more limited data package when specific information—such as non-clinical or clinical metabolite levels, or an assessment of API-derived impurities —is available. Each step in the flow chart corresponds to a dedicated section in the reflection paper.

Impurities that are also observed as metabolites at sufficient levels in animals and/or humans are discussed in Section 5.2. In such cases, the impurity may be considered qualified without the need for additional toxicological data. Section 5.3.2. addresses API-derived impurities that do not introduce new structural alerts compared to the API. These are generally considered to pose low toxicological concern and are assumed to be covered by the existing toxicological studies of the DS.

For all other impurities that do not fall within the categories described in Sections 5.2. and 5.3.2., the level of toxicological concern should be assessed as outlined in Section 5.4. This evaluation could consider factors such as the intended patient population, route of administration, and duration of treatment. If the outcome of this assessment indicates that the identified level of the NMI does not

pose a toxicological concern, no further qualification is considered necessary. In such cases, the impurity level may be considered qualified.

Figure 1: Flow chart of risk assessment for non-mutagenic impurities



If the level of toxicological concern warrants further evaluation, the assessment should begin with a literature search for existing impurity-specific in vitro or in vivo data relevant for a toxicological risk assessment. Where in vivo data is available, an Acceptable Level (AL) should be derived following the principles outlined in Section 5.5. In the absence of impurity-specific data, it should be determined whether suitable surrogate compounds with relevant toxicological data are available. If such surrogates are identified, read across should be conducted following the principles outlined in Section 5.3., and in vivo data for the surrogate may be used to derive an AL. If the level of the NMI is below the derived AL, the impurity is considered toxicologically qualified. If impurity level exceeds the AL, the level of concern should be re-evaluated (see Section 5.4.1.), taking into consideration the toxicological properties of the impurity and the extent to which the AL is exceeded.

In case impurity-specific data or data from a surrogate compound cannot be identified, New Approach Methodologies (NAMs) described in Section 5.6. may be applied to qualify the impurity. These include computational toxicology approaches (see Section 5.6.1.) e.g. using (Q)SAR models as well as in vitro approaches (see Section 5.6.2.) to evaluate the toxicological hazard of the impurity. NAM data may be combined in an integrated assessment approach, e.g. by building data into established AOPs for specific endpoints. Based on the predicted toxicity and the associated uncertainties of the computational models, a level of concern assessment should be conducted. If the potential risk is deemed acceptable, no further data are considered necessary. However, if the assessment indicates a

potentially unacceptable risk at the maximum daily intake of the impurity, the impurity level should be reduced, or additional toxicological data should be generated.

In such cases, in vitro approaches may be considered to further characterise the hazard of the impurity (see Section 5.6.2.). Only in exceptional circumstances—where none of the aforementioned methods have resulted in sufficient data to support qualification—should in vivo studies be considered to address remaining data gaps (see Section 5.7.).

5.2. Metabolites

In line with ICH Q3A(R2) and Q3B(R2), "impurities that are also significant metabolites present in animal and/or human studies are considered qualified". The term "significant metabolite" should not be understood in the context of ICH M3(R2), where it is defined as the threshold of 10% at which additional non-clinical characterisation of a metabolite is warranted. For the qualification of an NMI, existing non-clinical and clinical studies with the DS, in which sufficient exposure levels of metabolite have been demonstrated and the metabolite is chemically identical to the NMI, even when below 10% of the DS, are considered relevant to address the safety profile of the impurity. In this case the term "significant metabolite" is related to the extent of exposure level of the metabolite (E_M) in the (non-) clinical studies with the DS relative to the anticipated exposure level of the NMI (E_I) when being exposed to the maximum daily dose (MDD) of the medicinal product. As shown in the flow chart (Figure 1), the ratio (E_M/E_I) between the exposure to the metabolite and the anticipated exposure to the impurity should be established.

If sufficient levels of metabolite have been demonstrated and the metabolite is chemically identical to the impurity, the NMI can usually be qualified using the existing (non-)clinical studies with the DS to address the safety profile of the impurity.

While available animal metabolism data is normally used for qualifying an impurity, metabolism data in humans can also be used if the NMI is detected as a metabolite in humans. Generally, exposure levels of the metabolite should be reported based on metabolism data obtained in plasma. However, if scientifically justified, urinary metabolism data may also be acceptable to support the qualification of an impurity. The daily dose of an impurity is calculated as the percentage in the DP specification relative to the MDD. However, this does not consider the ADME properties to inform on the actual systemic exposure levels achieved for the impurity, complicating a direct comparison between levels of impurities and plasma metabolites for establishing the exposure margin. Weidolf et al. (2020) have proposed a pragmatic method for estimating the intrinsic exposure level of an impurity to compare with the observed level of plasma metabolite, thus establishing the E_M/E_I ratio. They propose to calculate the maximum theoretical concentration (MTC) achieved for the impurity (also referred to as E_I) and to compare this with the maximum observed concentration (MOC) of the metabolite (also referred to as E_M).

The MOC (or E_M) of the metabolite is taken from an animal toxicity study or a relevant clinical study. The average plasma concentration in the relevant dose group of animals or patients/volunteers would be considered to represent the MOC. Thus, E_M = average plasma concentration of the metabolite in animals or patients.

When calculating the MTC (or E_I), worst case assumptions are considered, such as complete bioavailability with no plasma-protein binding, no distribution into blood cells or other tissues, and no elimination. It is proposed to use the extracellular fluid (ECF) of the selected species as the minimal volume of distribution for estimating the systemic concentration of the impurity. In the paper by Weidolf et al. (2020), MTC is established based on the daily exposure of the impurity (mg/day), which

is calculated using the MDD and the impurity specification level, as well as an average rat ECF volume of 80.4 ml. Alternatively, if calculating the MTC to compare with human metabolism data, a human ECF volume of 14 L can be used (Tobias, 2025). Thus, E_I = MDD x %specification $level_{impurity}$ / ECF, where ECF in humans and rats is 14 L and 80.4 mL, respectively.

For a metabolite/impurity of low concern, no further non-clinical data is needed to qualify the impurity when unity is reached ($E_M/E_I \geq 1$). However, when potential toxicity of the metabolite/impurity is of concern, a larger E_M/E_I needs to be considered to ensure that unacceptable exposure due to the additional presence of the substance as a drug-related impurity is not achieved. A metabolite is considered of low concern when it has been identified at proportionate levels in animals and humans and it is not a human-specific or significant human metabolite (i.e. ≤ 10%). In case of a significant human metabolite, a disproportionate metabolite or a human-specific metabolite, it should be sufficiently characterised to justify that the metabolite does not significantly add to the toxicity profile of the parent before it can be considered of low concern. If exacerbated toxicities of the metabolite cannot be ruled out or if no toxicological data is available for the metabolite, an $E_M/E_I \geq 10$ should be applied.

5.3. Read across

If no sufficiently robust in vitro or in vivo toxicity data can be identified on the impurity itself, it is possible to perform read-across to one or more surrogate compounds for which robust in vitro or in vivo toxicity data is available. The single surrogate approach or the grouping approach can be employed to identify qualitative or quantitative in vitro or in vivo toxicity data for chemically similar compounds that can be used to qualify the impurity at the specified level. API-related impurities may also be assessed in terms of similarity between the API and the impurity to determine if the non-clinical experimental data for the DS can be used to qualify the impurity.

5.3.1. Surrogate approach

In a read-across framework, three different steps need to be considered: 1) Chemical similarity assessment between target and surrogate molecule, 2) Toxicological assessment of data and 3) Establishing an AL.

Chemical similarity assessment between target and surrogate molecule

When performing read-across to a surrogate compound, firstly, the impurity should be characterised in terms of chemical-structural and physicochemical properties as well as pharmacokinetic (PK) properties when available. Comparability based on physicochemical properties or PK properties, where available, should be discussed.

Relevant toxicophores that are present in the impurity should be identified, where a toxicophore is defined as a chemical structure or part of a structure that is related to the toxic properties of the impurity. This can include both pharmacologically active and non-active moieties of the impurity. Physicochemical properties (such as polarity, solubility, lipophilicity, ionisability, and molecular weight), as well as PK properties (such as bioavailability, distribution, metabolism, and excretion) should be considered, e.g. from databases or based on predictions using computational tools. When in silico tools are used, it should be justified that these are fit-for-purpose (see section 5.6.1.). Also, considerations regarding biological plausibility (e.g. mechanism/mode of action) may be included in the assessment.

Surrogate compounds, for which robust data is available, should be identified based on similarities to the NMI. The presence of the identified toxicophores of the impurity should be demonstrated in the surrogate compounds, and further, the presence of other functional groups – especially those spatially

close to the toxicophore, which could potentially affect the biological activity, should be identified. Multi-metric approaches (including Cosine, Tanimoto, and/or other similarity metrics), combined with the use of well-established molecular descriptors when appropriate, can be applied to assess molecular similarity (Sheridan et al., 2002 and Willett et al., 1998). The choice of adequate surrogate(s) should be justified based on the similarity and uncertainties with the read-across method and the adequacy of the outcome of the assessment should be provided together with the overall outcome of the read-across approach.

As detailed in the Computational toxicology section (Section 5.6.1.), different tools for predictions could be used for identifying toxicophores associated with endpoint-specific toxicities, e.g. (Quantitative) Structure–Activity Relationship (Q)SAR, as well as for predicting physicochemical and PK properties. It is acknowledged that the only endpoint in (Q)SAR modeling currently considered regulatory validated is mutagenicity in bacteria, as described in ICH M7(R2). Nonetheless, computational tools could be used for identifying toxicophores considered relevant for major targets (liver, kidney, cardiovascular system (CVS), gastrointestinal tract (GIT), skin, CNS and respiratory system) to support the read-across assessment. Further recommendations are given in Section 5.6.1. on the use and outcome reporting of (Q)SAR tools to predict toxicity and identify toxicophores.

Toxicological assessment of data

When one or more suitable surrogate compounds have been identified, an evaluation of the robustness of the available toxicological data should be performed to conclude whether the data is sufficient to establish an AL or perform a hazard assessment. Toxicity studies covering different endpoints could be used for informing on hazard or potency level, and it should be justified that the studies and endpoints for the surrogate compound are also relevant for establishing an AL or de-risking the impurity.

Establishing an AL

Based on the outcome of the read-across assessment, quantitative data on a surrogate could be used to derive an AL as defined in Section 5.5., while qualitative data could be used to de-risk an impurity as not adding significantly to the toxicity of the DS. If more than one surrogate is used to support a read-across assessment and ALs are calculated for each surrogate, the most conservative value should be used to set the AL for the NMI, unless there is convincing evidence that the impurity is less potent, and a higher AL could be accepted.

5.3.2. API-like vs. non-API-like impurities

All the degradation- and process-related impurities of the API that are structurally similar can be considered API-like and allow for read-across to the non-clinical studies with the DS for the purpose of qualification of the NMI. This may include degradation products of oxidation or hydrolysis reactions, where minor changes to the structure are introduced, and the difference does not affect the overall structure or size of the molecule. Specifically, the majority of chemical functional groups should be preserved in an API-like NMI while only 1-2 groups are changed via e.g. degradation by oxidation or hydrolysis. Physicochemical parameters should not be significantly impacted, and it should be justified that changes would not lead to increased toxic potency compared to the API. Moreover, it should be shown that the new chemical groups do not form a new toxicophore compared to groups present in the API. The term toxicophore is defined in section 5.3.1. Susceptible groups in the API includes e.g. carboxylate esters, amides, or carbamate groups that are prone to hydrolysis and carboxylic acids, carboxylate esters, hydroxyl groups, unsaturated hydrocarbons, amide groups and amine groups that are susceptible to degradation by oxidation (Xiao, 2022). Also, in cases where the impurity is a dimer or trimer of the parent structure, the impurity can be considered API-like if it can be justified that the dimerisation bridge does not introduce a new toxicophore (e.g. empagliflozin sugar dimer) and the

impurity will degrade back into the parent structure once it has entered the systemic circulation. If degradation of a dimer is not likely to take place, further qualification of the impurity should be considered in terms of altered physical-chemical properties and biological activity.

As scoped out in the section on read across and computational toxicology, similarity may be evaluated by using computational predictive tools that identify toxicophores and predict physicochemical and PK properties for the impurity, to compare the properties between the API and the impurity. This includes an evaluation of whether new toxicophores are introduced and whether physicochemical parameters are impacted, such as molecular weight, lipophilicity, and water solubility. When it is concluded that no new toxicophores are introduced in the NMI compared to the API, and that physicochemical and PK parameters are not significantly affected and therefore not likely to increase toxic potency compared to the API, the impurity is considered API-like. Thus, no further investigations are recommended, as the toxicological properties of the impurity are covered by existing non-clinical studies with the DS.

When impurities do show significant structural differences to the API, or where significant differences in physicochemical or PK properties are known and expected to affect the toxicity profile of the impurity, or where significant differences are predicted by computational tools, the impurity should be considered non-API-like and the next step in the flow chart should be followed.

Furthermore, there are very rare exceptions where a non-significant change of the chemical structure could considerably change the toxicological potency of a molecule (e.g. enantiomers, diastereomers). As in medicinal chemistry, such activity cliffs can be relevant for toxicity, especially where a key event is based on binding of the chemical entity to a specific site (Stumpfe et al., 2019), e.g. the S-enantiomer of thalidomide (Eriksson et al., 2000) exhibits teratogenic activity, while the R-enantiomer acts as a sedative; or the acid form of cholesterol-lowering statins that can condensate to the lactone form. These lactones can be present as impurities or formed endogenously. Despite the high structural similarity, the statin lactones may inhibit Complex III of the mitochondrial respiratory chain more potently and cause myopathy (Schirris et al., 2015). Cases where an API-like NMI with a non-significant change in relation to the API would result in increased toxicity is expected to be very rare. Identification of such potent NMIs should be based on identification of toxicophores or existing knowledge through available literature.

5.3.3. Grouping approach

Alternative to the single surrogate approach, a grouping approach can be used, where several similar impurities are grouped, containing the same toxicophores and functional groups, which allows the detection of trends across endpoints. Again, as defined for the read-across approach, adequate similarity must be demonstrated for chemical-structural, physicochemical, and PK properties to group the impurities. Based on the assessment of similarity of the grouped impurities, this may then allow for an overall AL to be derived for the group, which should be based on the most conservatively derived AL unless otherwise justified. Data from the grouping approach may also be used to de-risk the toxicological concern for the impurity, if it can be adequately demonstrated that the group of impurities is not likely to present any safety risk at the specified level of the impurity.

5.4. Level of concern considerations

The safety assessment of a non-mutagenic pharmaceutical impurity is a structured process, as depicted in the flowchart (Figure 1). An essential component of this process is the qualitative evaluation of the "level of concern," which relies on the factors presented in Figure 2. A detailed methodology, including examples, is provided in Annex 2. This tool is not intended to quantify risk, but rather to provide an integrated framework to guide decision-making at key stages.

The level of concern assessment is performed at two distinct points:

Step 1 – Initial screening

When an impurity cannot be qualified through straightforward approaches (e.g. existing guidelines, human metabolite status, structural similarity to the API), a preliminary level of concern assessment is conducted. At this stage, the factors in Figure 2 are used to determine whether further toxicological evaluation is warranted. By considering elements, such as the daily intake of the NMI, duration of treatment, clinical indication, and target population, it may be concluded that the NMI poses a sufficiently low concern to be considered qualified without additional data.

Step 2 – Risk Judgement

If further evaluation is needed, specific toxicological data are generated (e.g., determination of an Acceptable Limit (AL) or an in silico analysis). Significantly, if the impurity level is above the calculated AL, a second more comprehensive level of concern assessment should be performed. When computational toxicology is used to assess the impurity, a second level of concern assessment should be considered due to the associated uncertainties of the computational models. The objective of this second-tier assessment is to ensure the potential risk is acceptable within the overall context of the treatment. The acquired toxicological data are weighed against the full set of factors in Figure 2, with special attention given to the impurity's intrinsic properties (e.g., bioavailability, persistence) and the vulnerabilities of the patient population (e.g., children, patients with renal or hepatic impairment). This integrated analysis will determine whether the risk is acceptable, if the impurity limit must be reduced, or if additional studies are warranted. Ultimately, each risk factor must be evaluated in the context of all others to support a scientifically sound and balanced judgement.

5.4.1. Key Factors for Each Step

The following section outlines the priority factors from Figure 2 that should be considered during each stage of the level of concern assessment

Key Factors for Step 1 (Triage and Scoping)

At this stage, the objective is to make an initial decision on the need to proceed with further evaluation. The focus is therefore on the most impactful and well-defined factors of the analysis that do not require in depth evaluation to obtain the information.

Daily intake (of impurity): This is the primary starting point. A very low exposure (e.g., < TTC) can justify a rapid conclusion of the process.

Duration of treatment: A lifetime (chronic) treatment is of significantly higher concern than a short-term or single-dose ("once in a lifetime") treatment.

Clinical indication/severity of the condition: An impurity in a medicinal product for a life-threatening disease (e.g., chemotherapy) is viewed with greater tolerance than an impurity in a medicinal product for a benign condition.

Key Factors for Step 2 (Evaluation and Final Decision)

At this stage, more nuanced factors are integrated to refine the contextual understanding of the identified hazard that requires an acquisition of more product-specific information.

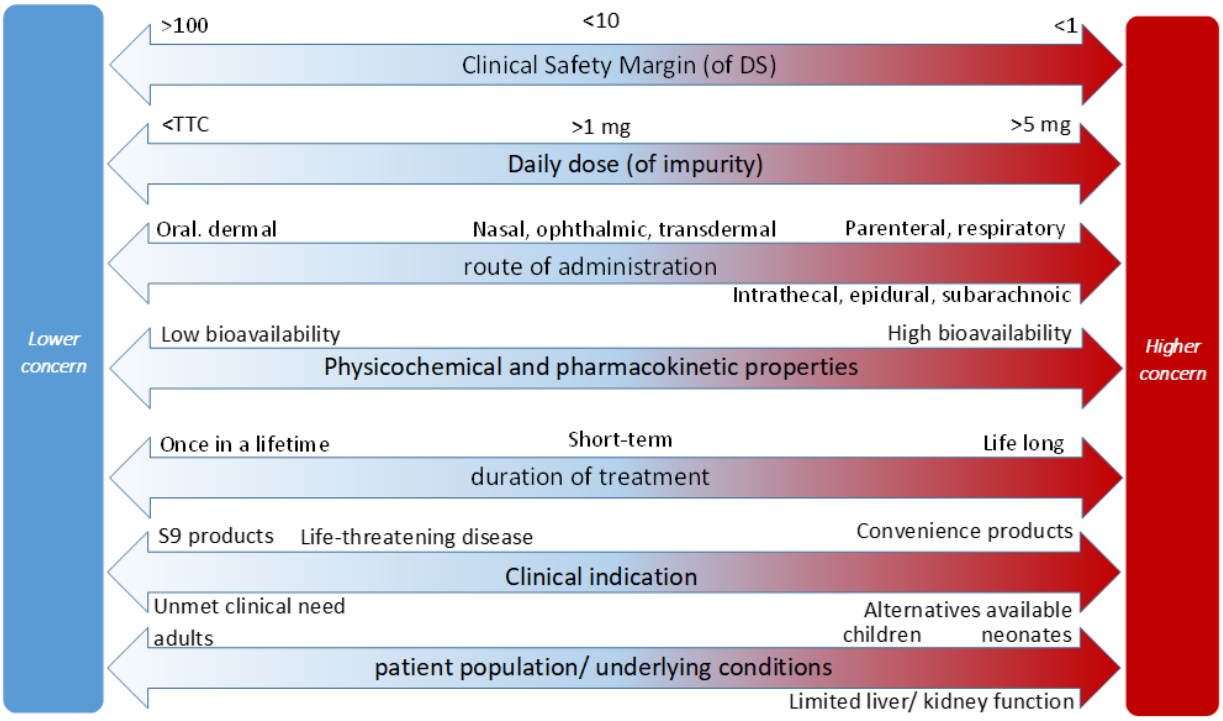
Clinical Safety Margin (of DS): Qualification is supported by the safety margin established during non-clinical testing of the DS. When the exposure levels of impurities in the toxicology batches significantly exceed the intended clinical exposure (large safety margin), and no adverse effects are observed at the NOAEL, the impurities are considered to have a low toxicological concern.

Physicochemical and pharmacokinetic properties: These properties could influence systemic exposure and organ-specific toxicity and are essential for interpreting both in vivo and in silico data and therefore crucial for interpreting the toxicological data. A persistent and highly bioavailable impurity is more concerning than one that is rapidly degraded and eliminated, even if their calculated AL was the same.

Route of administration: The route of administration (e.g., intrathecal) can significantly increase organ-specific exposure and thus the level of concern, even for a low systemic exposure.

Patient population (specific subpopulations): A limit deemed safe for general target patients may be inappropriate for vulnerable subpopulations, e.g. pediatric patients, patients with underlying conditions such as renal or hepatic impairment, who may exhibit reduced clearance and increased systemic exposure, pregnant women.

Figure 2: Aspects that can be considered to determine level of concern



5.4.2. Exposure level considerations

Clinical safety margin of the drug substance

When impurities are present in DS batches that have been evaluated in non-clinical safety studies, the NOAEL typically reflects the toxicological profile of the API, as impurities are not usually present at levels sufficient to drive toxicity unless specifically investigated. Nevertheless, a large safety margin—defined as the ratio between the exposure at the NOAEL in animal studies and the maximal anticipated clinical exposure—suggests low toxicity. This implies that both the API and any co-present impurities are unlikely to pose significant safety concerns at the tested levels. Thus, this information may support the toxicological qualification of impurities that were present in the test material used during (non-) clinical safety studies, provided their levels were representative of the intended clinical exposure.

Absolute daily intake of impurity

The absolute daily dose or exposure to an impurity is a critical factor in evaluating its toxicological risk. As illustrated in Figure 2, a daily dose of the NMI at or below the TTC is generally considered to represent a low concern. This concept will be explained below. In the regulatory context of pharmaceutical impurity assessments, a QT of 1 mg/day (or 0.15% whichever is lower) has been applied for DSs with a MDD of 667 mg/day or higher (ICH Q3A(R2)). Assuming a body weight of 50 kg, this corresponds to 0.02 mg/kg bw/day.

Recent publications (Graham et al., 2021; Harvey et al., 2017; Kenyon et al., 2024; Slikkerveer et al., 2024; Hasselgren et al., 2024) argue that APIs with a NOAEL below 0.02 mg/kg bw/day are extremely rare and that this would support the notion that 1 mg/day should be considered as a generally safe threshold for NMIs. For chemicals of low toxicological concern, exposures below this level of 1 mg/day could indeed be acceptable in most cases. However, this assumption cannot be applied universally. Certain chemical classes and chemotypes – such as organophosphates, carbamates or beta-lactams – are known to exhibit toxicity at lower exposure levels (Harvey et al., 2017). Therefore, safety concerns cannot be excluded a priori. Additional considerations, such as inter- and intraspecies differences in sensitivity, differences in bioavailability and duration of exposure – animal studies are often short-term, whereas patients may receive treatments over extensive periods – may affect the risk assessment of an NMI. Although many of these factors have been considered by Kenyon et al. (2024), EMA does not regard the 1 mg/day threshold as a definitive cut-off value. Instead, it is treated as a reference point - the midpoint on the daily dose scale illustrated in Figure 2. Thus, when impurity exposure levels exceed the TTC but remain below 1 mg/day, the level of concern will have to be evaluated, taking into account all relevant factors outlined in Figure 2. If the daily exposure to the NMI exceeds 1 mg/day, the level of concern increases proportionally, warranting more detailed toxicological assessments.

Table 1: Daily impurity exposure threshold

Daily impurity exposure threshold	Associated level of concern	Rationale and recommended action
≤ TTC	Low	Exposure is of low concern.
> TTC and < 1 mg/day	To Be Assessed	The level of concern must be evaluated on a case-by-case basis, considering all other risk factors (as shown in Figure 2).
≥ 1 mg/day[1]	Increased to High	Exposure at or above 1 mg increases the concern to high.

[1]Derived from the ICH Q3A guideline as a QT. Recent publications (Graham et al., 2021; Harvey et al., 2017; Kenyon et al., 2024; Slikkerveer et al., 2024) support it as a safe limit for most impurities, but it is not a definitive cut-off value.

Threshold of Toxicological Concern

The TTC is a risk assessment tool used to evaluate low-level exposure to chemicals with limited toxicological data. As described in ICH M7(R2), the TTC is well-established for mutagenic impurities. For non-mutagenic endpoints, the same principle can be applied when insufficient toxicological data are available and read-across is not feasible (see section 5.3.). Since TTC values are endpoint- and route-specific, their applicability must be verified for each case. If the exposure level is below the relevant TTC, no further action is generally required. However, TTC values represent thresholds below which no appreciable risk is expected for most—but not all—chemicals. Therefore, even when exposure is below the TTC, the level of concern must be evaluated in a broader context (see Figure 2).

The international regulatory community has established specific systemic and local toxicity thresholds for extractables and leachables in oral and parenteral formulations. These values are based on a large dataset establishing PDEs for impurities of different potencies, resulting in robust thresholds divided across different durations of administration and these values may inform decision making in evaluating the safety risk of an impurity. Thus, if NMIs in oral or parental medicinal products are below the systemic or local toxicity thresholds in Table 2, they are in general considered to be of low concern and no further data is considered necessary to qualify the impurities. Assessment of higher potency impurities should be based on a review of available impurity-specific data or computational toxicology alerts based on the presence of toxicophores in the NMI.

Table 2: Systemic and local toxicity thresholds for non-mutagenicity endpoints

Systemic Toxicity Thresholds

Exposure Duration	Oral	Parenteral, Dermal/Transdermal, Inhalation
> 1 month	48 µg/day	12 µg/day
≤ 1 month	136 µg/day	26 µg/day

Local Toxicity Thresholds

Topical Ophthalmic	Subcutaneous and Intradermal	Dermal and Transdermal	Intracerebral, Intrathecal, Epidural and Intraocular	Inhalation
20 ppm	50 ppm	500 ppm	Impurity-specific evaluation	5 µg/day

Toxicity threshold values for inhalation and dermal/transdermal routes have been established based upon parenteral thresholds in lieu of available PDE values. These thresholds are currently being discussed in the international regulatory community for extractables and leachables and are considered relevant for other non-mutagenic impurities.

Within the food safety domain, TTC values have been established for various chemical classes. For organophosphates and carbamates, the relevant TTC is 0.3 µg/kg bw/day. All other chemicals are grouped according to the Cramer classification, with TTC values of: Cramer Class I: 30 µg/kg bw/day; Cramer Class II: 9 µg/kg bw/day and Cramer Class III: 1.5 µg/kg bw/day (EFSA, 2019; Munro et al., 1996). These values are applicable to NMIs in orally administered pharmaceuticals, provided that the impurity is not exempted from the application of the TTC (e.g. steroids, bio-accumulative substances, or those with known high potency). Alternative TTC values derived using refined methodologies (e.g. Tluczkiewicz et al., 2011) are generally consistent with the above. Future refinements of the classification of chemicals may emerge from initiatives such as the Food and Drug Administration's Expanded Decision Tree (EDT) Project, which aims to modernise the Cramer classification and improve TTC applicability (Stice and Adams, 2025).

More work has been done to derive TTC values for other routes of exposure such as inhalation (e.g. Escher et al., 2010; Tluczkiewicz et al., 2016; Nelms and Patlewicz, 2020). Cramer classification appeared to be a less suitable approach for this route. For the chemicals that were grouped as toxic or reactive, TTC values in the range of 2-4 µg/day have been calculated. This compares reasonably well with the QT of 5 µg/day for leachables in orally-inhaled, and nasal DPs as derived by the PQRI consortium (Ball et al., 2007). The latter mentions that irritating chemicals, including aldehydes, nitriles and isocyanates, as well as metals and metal salts, could be of concern below this value. A TTC of 4 µg/day can be considered as a practical value for impurities in medicinal products administered via inhalation, provided they do not belong to one of the chemical classes of high concern.

For the dermal route of administration, skin sensitisation is considered the most sensitive non-mutagenic endpoint when it concerns reactive chemicals, including high potency category (HPC) chemicals (Roberts et al., 2015; Nishijo et al., 2020). Based on an extensive review of literature data, dermal sensitisation thresholds (DST) have been calculated for non-reactive, reactive and HPC chemicals, which are 900, 64, and 1.5 µg/cm^2, respectively (Safford et al. 2008, Safford et al. 2015; Nishijo et al. 2020, Parris et al., 2024). These were modified later to 710, 73, and 1.0 µg/cm^2, respectively (Chilton et al., 2022). These values could be used for dermal products, to address the concern for dermal sensitisation. For other non-mutagenic endpoints, systemic exposure needs to be considered, taking into account the surface to which the medicinal product is applied and the degree of dermal absorption (see below for discussion on PK properties and bioavailability). Once the systemic exposure has been estimated, a comparison with parenteral TTC values can be made.

Intramuscular, subcutaneous, and intravenous routes of administration are the main parenteral routes for pharmaceuticals. Where parenteral is discussed here, any of these three routes are considered. For other parenteral routes, specific considerations may apply as the pharmaceutical may be administered into a small compartment or in close contact with sensitive tissue (e.g. ophthalmic products or intrathecally, epidural, or sub-arachnoidally administered products). For any route of administration not discussed in this reflection paper, like intravitreal route of administration, a case-by-case discussion would be needed. In the literature, reports have been published describing various approaches to derive parenteral (systemic, internal) TTC values. Arnot et al. (2022) used the Munro database to derive internal NOELs (mg/kg/day) by combining the oral NOEL values with available PK data, where such data were not available, by applying Physiologically Based Pharmacokinetic (PBPK) modelling to estimate internal NOELs. Internal TTC (iTTC) values were derived using the fifth percentile as cut-off and dividing this by 100. For whole body, an iTTC value of 0.5 nmol/kg was proposed. Taking a human body weight of 50 kg and using the median molecular weight (220) of the compounds in the Munro dataset, this can be transformed to 5.5 µg/day. This value is remarkably similar to the one that was derived by Partosch et al. (2015), who used different databases and arrived at an iTTC for Cramer class II/III compounds of 0.1 µg/kg bw/day or 5 µg/day for a person of 50 kg. In ICH Q3D modifying factors have been proposed allowing the derivation of a parenteral PDE from an oral PDE (mg/day) taking into consideration the oral bioavailability. In worst cases, where oral bioavailability data are not available, a parenteral PDE can be extrapolated by dividing the oral PDE with a modifying factor of 100. For most of the elemental impurities, parenteral PDEs have been determined by dividing the oral PDE by a factor of 10, which assumes that estimating oral bioavailability at 10% is sufficiently conservative. If the modifying factor of 10 were to be applied to the Cramer class TTCs, we would arrive at systemic TTC values of 150, 45 and 7.5 µg/day for a 50 kg person for Cramer class I, II, and III compounds, respectively. In this approach, the systemic TTC for class III compounds is quite close to the iTTC values proposed by Arnot et al. (2022) and Partosch et al. (2015). These values differ, however, from the lifetime parenteral TTC value of 35 µg/day that took into consideration the 422 compounds in the Extractables and Leachables Safety Information Exchange (ELSIE) database. This database contains toxicity data for reported or presumed extractables and leachables (Masuda-Herrera et al., 2022). In this publication, corrections for bioavailability were based either on actual PK data or

on an in silico tool for estimating bioavailability. If no NOAEL was available, the LOAEL was chosen, and an additional correction factor was used. The estimated systemic values were divided by 100.

In addition to the threshold values described above, the route-dependent differences in toxicity must be specifically considered. These route-specific sensitivities are also reflected in the different TTC values; a principle now discussed in relation to extractables and leachables, which proposes distinct systemic and local toxicity thresholds based on the route of administration. For instance, orally-inhaled and nasal DPs are delivered to the respiratory tract where tissues are receptive to sensitisation and irritation, a risk acknowledged in guidelines such as ICH S4, which covers inhalation toxicity studies. Thus, these endpoints are often the most critical for this route.

When dermally applied, sensitisation is the most sensitive endpoint when HPC chemicals are concerned, and possibly also when the impurity is a non-HPC reactive chemical. The concept of dermal sensitisation threshold (DST) is a key tool for this assessment. However, given the possibility of dermal absorption, for the dermal route systemic toxicity should be considered as well. For some specific routes of administration into small confined spaces such as intrathecal, epidural or sub-arachnoidal, the relatively high local concentration is an additional risk factor. In these situations, an estimate of the local concentration would be a better parameter for evaluation than the daily dose. In addition, the high sensitivity of central nervous system (CNS) tissues needs to be considered for these special routes.

Table 3: DST and TTC values for non-mutagenic endpoints

Route of administration	**Class**	**DST[1]**	**TTC[2]**
Oral	Cramer class 1		1500
	Cramer class 2		450
	Cramer class 3		75
	organophosphates or carbamates		15
Orally-inhaled or nasal			4
Dermal	Non-reactive[3]	710	
	Reactive (non-HPC)[3]	73	
	HPC[3]	1	
			Parenteral TTC/absorption[4]
Parenteral			5

1 DST (µg/cm2), relevant only for sensitisation as an endpoint.[2] Threshold of Toxicological Concern for non-mutagenic endpoint (µg/day calculated for a 50 kg person).[3] Classification according to Roberts et al. (2015). [4] for other non-mutagenic endpoints. HPC = High Potency Category

Physicochemical and pharmacokinetic properties / bioavailability of the impurity

By definition, medicinal products administered via intravenous injection have 100% bioavailability. Due to little or no metabolism in the skin or muscle, most medicinal products show between 60 and 100% bioavailability after subcutaneous and intramuscular injections (Stielow et al., 2023). Thus, parenteral routes pose the highest concern, as opposed to medicinal products administered via routes where limited absorption may reduce the systemic exposure. This is a foundational PK principle applied across all medicinal product development. Clearly, this is only relevant with respect to systemic toxicity. As discussed above, local toxicity is to be considered separately.

Information on absorption and bioavailability of the impurity may be retrieved from the literature. In the absence of such data, physicochemical (PC) properties can be considered to estimate

bioavailability. These properties are also used in in silico tools to estimate bioavailability. The use of such predictive tools is strongly endorsed by modern regulatory approaches, most notably in ICH M7(R2) for mutagenicity assessment. As these tools have their limitations, predictability can be improved by supporting experimental NAM data such as transport across Caco-2 cells and metabolism in hepatic models, a strategy encouraged by agencies like the FDA in its Predictive Toxicology Roadmap. In case the impurity exists as (dia)stereoisomers, potential effects on PK properties need to be considered. In the absence of factual data to the contrary, bioavailability of impurities in medicinal products administered via the respiratory route is considered to be (close to) 100%.

Impurities that are poorly degraded or eliminated otherwise increase the level of concern as such impurities can accumulate and, even with low daily exposures, may reach tissue concentrations where adverse effects could occur. Examples are highly bio accumulative impurities, where even short-term dosing can result in long term exposure, as is the case for example for certain per- and polyfluoroalkyl substances (PFAS). This potential for accumulation is a key reason for conducting the repeated-dose toxicity studies outlined in ICH M3(R2).

5.4.3. Clinical considerations

In the level of concern analysis, a case-by-case approach that considers the specifics of the target population and therapeutic indication is essential to define appropriate specification limits for impurities. This should also take into consideration the duration of treatment.

Duration of treatment

The treatment duration is a key factor to consider when determining the level of concern. For short-term treatments, the level of concern is usually lower, and the principle of applying higher QTs for shorter exposure durations is well-established and consistent with approaches in several ICH impurity guidelines. For example, the core guidelines for non-mutagenic organic impurities, ICH Q3A(R2) and ICH Q3B(R2), set QTs primarily designed for lifetime exposure. Furthermore, other guidelines like ICH Q3C(R9) (for residual solvents) and ICH Q3D (for elemental impurities) both indicate that it may be acceptable to exceed their recommended limits for short-term dosing (i.e., ≤ 30 days). Recommended threshold values in Table 2 take administration of shorter duration into account. Conversely, chronic treatments, particularly those that last throughout a patient's lifetime, necessitate a more thorough evaluation of impurity levels due to increased cumulative exposure. Therefore, when treatment duration increases, also the level of concern will increase.

Clinical indication

The clinical indication should be considered as a critical factor, as stipulated in the ICH Q3A and Q3B guidelines. These guidelines provide the primary criteria for qualifying impurities but explicitly permit changes to the QTs based on a scientific rationale that includes the clinical context.

In the context of severe or life-threatening diseases, the presence of impurities may be justified due to a different benefit-risk balance. This is most clearly articulated in the ICH S9 guideline, which specifically addresses the management of impurities in anti-cancer medications, noting that applying the same controls as for less severe conditions is inappropriate. This principle is a central component of the benefit-risk assessment frameworks used by major regulatory authorities. Furthermore, alterations in the clinical applications of marketed products may necessitate the re-assessment of existing impurity specifications.

Target population

The target populations should be considered in establishing the level of concern. The sensitivity to toxic effects can vary considerably among different groups. Key points to consider are:

- Paediatric populations: Children are often more susceptible to the toxic effects of impurities due to developmental differences in ADME (Absorption, Distribution, Metabolism, and Excretion). The specific needs and vulnerabilities of this population are detailed in ICH E11(R1), Clinical Investigation of Medicinal Products in the Paediatric Population.
- Patients with renal or hepatic disease: Impaired elimination can lead to increased concentrations of impurities. EMA have published specific guidance documents requiring PK assessments in these populations to ensure safety.
- Pregnant individuals: Exposure to impurities during pregnancy is concerning due to potential developmental toxicity. The nonclinical assessment of such risks is a core topic of the ICH S5(R3) guideline on reproductive toxicology.

5.5. Acceptable Level calculation

Impurity-specific data and data on a surrogate considered acceptable for applying read-across to the impurity can be used for deriving ALs for an NMI. Impurity or surrogate specific toxicological data should be of high quality in vitro and/or in vivo studies. Deriving an estimate of an AL of exposure for a new NMI in patients by using toxicological data is generally achieved by choosing an endpoint specific Benchmark Response (BMR) in an acceptable study and calculating the Benchmark Dose Lower boundary (BMDL) as a point of departure (PoD). Alternatively, an experimentally defined NOAEL may be used as a PoD, but generally provides a less accurate PoD. The PoD is the anchor point for the AL calculation and applying AFs to correct for variability, uncertainties and known differences between the animal model used and the patient for whom a predicted safe level is needed. Guidance and recommendation for identifying relevant BMR can be found in e.g. WHO EHC 240: Principles for Risk Assessment of Chemicals in Food; Chapter 5 Dose-response assessment and derivation of health-based guidance values (second edition, 2020) or EPA Benchmark Dose Technical Guidance (EPA/100/R-12/001 – June 2012). This reflection paper proposes the AL method, by which similar toxicological principles are used, e.g. as described in ICH Q3C and Q3D. The AL is considered a conservative estimate of the dose level below which patient safety is not affected and represents the maximal daily dose of an impurity in a pharmaceutical product to which a patient can be exposed during treatment without compromising the patient's health. Although the methodology is similar to the methods for deriving a PDE, the PDE is a generally applicable safe level of exposure for a specific impurity in any product, whereas the AL is a level for an impurity that has been set for a specific product and is considered acceptable in that context. Therefore, specific considerations regarding the level of concern relevant to the product can also be considered (see below). The methodology starts with the selection of the PoD, which can be a BMDL, a NOAEL, or, in the absence of these, a Lowest Observed Adverse Effect Level (LOAEL) or any other acceptable PoD. Subsequently, AFs are applied for inter- and intraspecies variability (AF1 and AF2, respectively), duration of the study from which the PoD is taken (AF3), severity of the toxicity (AF4) and the absence of a BMDL or NOAEL (AF5). Furthermore, a body weight that is appropriate and conservative for the target population (e.g. pediatric patients) should be selected when calculating an AL.

Impurities cover a wide chemical space and thus, bioavailability via different administration routes may vary greatly. Often toxicity data are only available for a single route, mostly the oral route, whereas pharmaceuticals are administered via other routes, e.g. a parenteral route. It is therefore

recommended that when route-to-route extrapolation is needed, an additional AF is used to account for differences in bioavailability (AF6).

In this reflection paper, the use of read-across is described as an alternative when insufficient impurity-specific toxicity data are available. To account for the additional uncertainty that may be introduced by relying on the toxicity data of a surrogate, the use of another AF is recommended (AF7).

The AL can be calculated with the formula:

$$AL\ (\frac{\mu g}{d}) = \frac{\text{PoD}\ \left(\frac{mg}{kg}/d\right) \text{ x } 50\ \text{kg x } 1000}{\text{AF1 x AF2 x AF3 x AF4 x AF5 x AF6 x AF7}}$$

The use of AFs is described in more detail in the Annex.

The most relevant study should be used to select the PoD, taking into consideration the duration of human exposure, the duration of the animal study, the species used, the route of exposure, the toxicological endpoints monitored, and the quality of the study data.

It may happen that an effect is observed that is not relevant for humans. In that case it could be inappropriate to choose this endpoint as the basis for the PoD and another one should be considered, rather than establish an AL on the basis of this endpoint not considered relevant for humans.

If it is unclear what is the most appropriate PoD, it is acceptable to calculate multiple AL values and select the most conservative value. It is not recommended to utilise LD_{50} values in AL calculations.

BMDL can be used as PoD. A BMDL makes use of all data on the dose-response curve and is the preferred option from a scientific point of view. When a BMDL is used, the proper derivation of this BMDL should be established, considering crucial elements such as the choice of the critical effect size, the number of dose groups, the BMD credible interval (BMDL-BMDU) and the ratio between the BMDL and the lowest dose. Further considerations for deriving a BMDL from experimental animal studies is given in the section on in vivo qualification studies. If no reliable BMDL can be derived, a NOAEL can be used as PoD. When no NOAEL has been established, a LOAEL can be used as PoD.

The AL method is distinctive from the PDE method described in ICH Q3C and ICH Q3D, as it is meant to derive a product-specific limit for an impurity and not aimed at setting an authorised limit generally applicable for all products. This allows for a case-by-case approach that considers product-specific aspects and takes the body weight of the target population into account. Furthermore, the AL method includes corrections for bioavailability (as also done in Q3D) and considers uncertainty related to a surrogate approach, whereas the PDE method does not.

5.6. New approach methodologies

New Approach Methodologies include in silico, *in chemico* and in vitro approaches making use of existing data or applying non-animal models to characterise the hazard or risk of a compound. It is still an area under development with many stakeholders being active in developing new approach methodologies as an alternative to animal models, such as Animal-free Safety Assessment of Chemicals: Project Cluster for Implementation of Novel Strategies (ASPIS) (RISK-HUNT3R, ONTOX and PrecisionTox) and EPAA (European Partnership for Alternative Approaches).

5.6.1. Computational toxicology

If impurity-specific data or the read-across approach have not resulted in relevant data for qualifying the impurity, computational predictive in silico tools comprising of one or more endpoint-specific models can be used to identify potential safety alerts i.e., toxicophores of the impurity, or to further characterise an identified safety concern (see section 5.6.4.). Development of predictive models for adverse reactions has been described in literature and could provide a major advantage in reducing the need for animal testing to qualify an impurity (Bender et al., 2007).

Computational toxicology refers to the use of computational, in silico methods to predict the potential toxicity of compounds without the need for traditional animal testing. This includes predictions of (Quantitative) Structure-Activity Relationships ((Q)SAR) that is usually based on trend analyses across broad databases and training sets of chemical compounds tested in vitro and in/ex vivo, which can be processed using Artificial Intelligence (AI) and Machine Learning (ML) to analyze the data (Ajisafe et al*.* 2025).

Validation of the model

When an in silico model is used in the risk assessment of an impurity, the validity of the model for the specific prediction should be justified by the Applicant. The model should follow the five validation principles for (Q)SAR models according to the Organisation for Economic Cooperation and Development (OECD), that is, the model should have 1) a defined endpoint; 2) an unambiguous algorithm; 3) a defined domain of applicability; 4) appropriate measures of goodness-of-fit, robustness and predictivity; and 5) a mechanistic interpretation, if possible (OECD, 2004). Other important guidance includes the OECD Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship Model (OECD, 2014) and the OECD (Q)SAR Assessment Framework (OECD, 2023). The validity assessment should summarise the applicability of the model for the intended purpose, considering whether the pre-specified criteria for performance metrics (e.g. Matthew's correlation coefficient for binary predictors, MCC) and data interpretation are met, also defining potential limitations. In general, the key points for validating in silico prediction tools include, but are not limited to:

- Cross-validation (appropriate cross-validation techniques, e.g. k-fold cross-validation)
- Performance metrics (appropriate performance metrics relevant to the specific prediction task, such as accuracy, precision, ROC-AUC, MCC etc.)
- Benchmarking (comparison of the performance of the tool against existing methods or benchmarks to demonstrate its efficacy)
- Interpretability (interpretable predictions for intended use)
- Validation datasets (high-quality validation datasets that are representative of the problem domain and cover a wide range of scenarios)

Several best practice papers for validation of (Q)SAR methods have also been published, e.g. in a recent paper by Ajisafe et al*.,* which includes considerations of cross-validation, external validation, and benchmarking against traditional methods, as well as how to ensure model robustness and generalisability for ML/AI developed models (Ajisafe et al*.* 2025, Raies and Bajic 2016; Myatt et al*.* 2017, Worth et al., 2011).

It is the responsibility of the software provider to make the validation report for the software tool available with specific descriptions of the above parameters, defining the applicability domain of the model and performance within the domain. The model validity, applicability and adequacy should be

reported in the QSAR Model Reporting Format (QMRF). It is meanwhile the responsibility of the Applicant applying the tool to provide an expert review of the validity of the performed prediction, including that the target molecule is within the domain of applicability, which is typically reported in the QSAR Prediction Reporting Format (QPRF) (OECD, 2024). A summary of the validation assessment, e.g. in the form of a QPRF, should be submitted as an annex to the full in silico prediction report, including an expert review of the data, to conclude on the outcome and justify the validity of the prediction.

It is recommended to use the latest updated version of the tool available and to provide a full report of the in silico analysis in the submission. Of note, earlier versions may be justified (e.g. the tool has not undergone significant changes that affect prediction performance). Where available, the use of two complementary methods is recommended to enhance confidence in the prediction (statistical-based and expert rule-based). The absence of the complementary method for the chosen endpoint(s) should be justified.

Quantitative Structure Activity Relationship (QSAR) tools to predict potential toxicophores

In a first step, (Q)SAR tools can be used to identify potential safety concerns associated with the chemical structure or a portion of a structure (e.g., a functional group) of the impurity (toxicophores). The endpoints included should be able to define general (chronic) toxicity on major target organs and systems (liver, kidney, CVS, GIT, skin, CNS and RS), ideally in terms of functional and tissue organ changes. An alert for non-mutagenic carcinogenicity should also be considered and human relevance for mode of action should be addressed, e.g. via published literature and available tools/ databases to qualify the impurity. Any additional alert, e.g. for reproductive organs, reported by the chosen tool and not described among endpoints listed above, should be included in the expert review and its relevance considered in the overall level of concern. QSAR model predictions are most reliable if they come from the model's applicability domain. In case of an out-of-applicability-domain prediction for the impurity at the predefined endpoint(s), the prediction is not considered as reliable and an elaborated expert judgement with additional supporting evidence or alternative tools with a more suitable training dataset would be needed. For products administered topically on skin, a sensitising potential should be considered. Multiple predictive tools (commercial or free) are available for assessing e.g. general toxicity endpoints or skin sensitisation potential (reviewed by Teubner et al., 2013; Dik et al., 2014; Golden et al., 2020; Ta et al., 2021; Rim, 2020, Wei et al., 2024). It is also encouraged to use available open data sources and dashboards that enable access to collections of chemical hazard and risk information from public and governmental databases (Williams et al., 2017) as well as historical toxicological and pharmacological databases (Watford et al., 2019, Wang et al., 2017) to further determine whether the impurity has known toxicophores. Models for physicochemical and PK properties can be used to further define risks associated with the impurity along with identified toxicophores, e.g. in terms of anticipated absorption of the impurity and the potential impact on toxicity.

As stated above, the outcome of a prediction should be supported by an expert review of the data as well as an assessment of the validity of the outcome. While there is no adopted protocol for performing an in silico hazard assessment, several papers have been published by international consortia, including industry and regulatory experts, in an effort to develop in silico toxicological protocols to support hazard assessment of different target organ toxicities. Myatt et al. proposes a checklist for elements to consider as part of an expert review of a QSAR model result as well as how to report the outcome (Myatt et al., 2018) while Cayley et al. reviews common arguments used to resolve prediction scenarios made by complementary (Q)SAR models as part of the ICH M7 framework (Cayley et al. 2023). Additional literature addresses expert review of predictions for genetic toxicity (Hasselgren et al., 2019), liver toxicity (Bassan et al., 2021a), heart, kidney and lung toxicity (Bassan et al., 2021b) and carcinogenicity (Tice et al., 2021). These publications cover important points to consider when

using in silico tools to predict organ toxicity and performing an expert review of the outcome. Overall, key aspects of an expert review include a) Resolving conflicts, when different in silico models give different predictions (e.g., positive vs. negative) to decide on a final conclusion; b) Validating positive and negative predictions to ensure they are not false positives/negative by investigating supporting evidence for the prediction's limitations; c) Interpreting ambiguous results to clarify inconclusive or equivocal predictions; d) Consider the scientific rationale to justify the final conclusion by considering the models' training sets, methodologies, and the specific structural features of the impurity being evaluated; e) Ensuring applicability to the impurity in question by checking if similar structures or mechanisms were considered during model development.

5.6.2. In vitro approaches

In vitro models can be helpful to fill data gaps, *e.g.* in vitro models for transport and metabolism can strengthen the predictivity of in silico tools for bioavailability (Paixão et al., 2012; Schneckener et al., 2019) or they can be used to compare the potency of impurities for a specific in vitro endpoint (Escher et al., 2022; Rovida et al., 2021).

When (Q)SAR predictions raise concerns, further qualification data may be needed. Targeted use of in vitro methods (2D and 3D cell systems and micro-physiological systems) with careful selection of endpoints may be considered. Currently, no single assay would provide a definitive answer to the question whether an impurity can be considered safe at the specified level. Scientific efforts are ongoing to develop batteries and strategies for using in vitro approaches. When applying an in vitro approach to evaluate the safety of a NMI, assays should be carefully selected based on concerns identified from SAR or read-across analyses and their applicability justified. Targeted in vitro models might not be validated for their use for regulatory purposes. This should not prevent the use of non-standard in vitro methods. To facilitate an assessment of the quality of data produced and their potential utility in regulatory applications, supportive information should be provided, showing that the method is suitable for its intended purpose. Useful guidance to this end can be found in the Guideline on the principles of regulatory acceptance of 3Rs (replacement, reduction, refinement) testing approaches (EMA/CHMP/CVMP/JEG-3Rs/450091/2012) currently under revision and the OECD Guidance document for describing non-guideline in vitro test methods (OECD 2014).

It is expected that more valuable information would be obtained from assays in which the purified NMI is tested. Adding spiked samples of the DS to test systems would complicate the interpretation of the read-outs, as the DS itself may also have an effect in the in vitro model employed.

Several papers have been published in which industry consortia investigate and discuss secondary pharmacology in vitro screening panels for detecting off-target activity at relevant receptors and enzymes that are known to be associated with toxicities, such as different kinases, hERG, nuclear hormone receptors and G-protein coupled receptors. Specific targets for in vitro screenings are recommended in these papers to cover a broad range of organ toxicities in screening panels already used by industry in medicinal product development and lead optimisation (Bowes et al., 2012, Lynch et al., 2017, Blum et al., 2023, Brennan et al., 2024). The proposed targets in the papers, especially the more recent paper from Brennan et al., 2024, could form a suitable strategy for further in vitro testing to demonstrate lack of concern for associated toxicities for the NMI.

5.6.3. Adverse Outcome Pathways (AOPs)

The AOP methodology is an approach that provides a framework to collect, organise and evaluate relevant information on biological and toxicological effects of chemicals. More specifically, the AOP approach organises existing knowledge concerning biologically plausible and empirically supported links between molecular-level perturbation of a biological system and an adverse outcome at a level of

biological organisation of regulatory concern (OECD, 2017). An AOP can be used to combine evidence from several in silico and/or in vitro studies to fill in knowledge gaps regarding toxicological events. OECD has developed guidance on the development and assessment of AOPs as well as an AOP framework for organising data at the chemical and biological levels (OECD, 2017) where several AOPs addressing different targets have been included in the AOP Wiki.

If sufficiently justified, AOPs could be used to both qualitatively or quantitatively to qualify an NMI by either documenting the lack of hazard of a compound for major organ toxicities, or to derive a NOAEL or BMDL for setting an AL.

5.6.4. Hazard characterisation and quantitative risk estimation

NAM primarily provide qualitative data, and their use in quantitative risk estimation remains limited. To enable quantitative interpretation — particularly, the translation of quantitative in vitro data to the in vivo situation (QIVIVE) — additional data and methodological development are needed. As a result, the current application of NAM tools is often focused on hazard identification and characterisation (Schmeisser et al., 2023). When NAM tools indicate the absence of relevant hazards for major targets (liver, kidney, CVS, GIT, CNS and respiratory system), this information can be incorporated in a weight-of-evidence approach for the safety assessment of the NMI. Conversely, if a potential hazard is identified using NAM tool, it must be demonstrated that the potency of the impurity to elicit the associated toxicity is not of concern. In such cases, it may be sufficient to justify that the exposure to the impurity at the proposed specification limit is without safety concerns based on a weight of evidence assessment of toxicity and PK considerations.

5.7. In vivo qualification studies

5.7.1. Design of in vivo studies

Conducting additional in vivo studies to qualify the toxicological properties of new impurities, as suggested in ICH Q3A(R2), is discouraged. This is in line with the 3Rs principles and the limited scientific relevance and reliability of such studies in this context. In particular, in vivo studies using impurity-spiked batches of the DS at dose levels at or below the NOAEL of the DS have historically provided limited additional toxicological insight. However, if a new NMI cannot be controlled below DST/TTC thresholds, and all other alternative qualification approaches mentioned above have been exhausted, an in vivo study may be considered. In such cases, a preferred study design is proposed below to ensure harmonisation of the approach for deriving a PoD for setting an AL based on in vivo data.

Several industry-led publications have investigated the preferred study design and have provided recommendations for an in vivo study design (Mitra et al., 2021; Slikkerveer et al., 2024). While some of the principles from these publications are endorsed, others require further reflection, particularly regarding ensuring sufficient exposure to the impurity to establish an adequate exposure margin relative to the proposed specification limit of the impurity. As noted in both papers, the commonly used test item is a DS batch spiked with impurity. However, this design is flawed as the impurity levels may not be sufficient to ensure an adequate ratio between the impurity exposure and the AL. This uncertainty may lead to repetition of in vivo studies, which is undesirable from both scientific and ethical perspectives. Therefore, it is recommended to conduct the in vivo study using neat samples of the impurity (i.e. isolated impurity with a purity of > 95%), allowing for direct evaluation of the effects of the impurity itself without confounding influences from the DS. Moreover, the in vivo study should be GLP compliant, conducted following OECD test guideline 407, designed for 28 days of repeated

dosing generally in rats via the clinical route of administration. For medicinal products intended for short term use, the duration of the in vivo study could be reduced to 14 days. No recovery period for the treated groups is needed, but a vehicle control group should be included. Finally, toxicokinetics (TK) should be included but maybe integrated in the main study e.g. as part of the high dose group, avoiding the need for a separate TK group.

Another issue is the dose selection and number of animals per group. The overall recommendation from the papers is to use five animals/sex/group, including a high and a low dose level, to allow for deriving a NOAEL as PoD. The BMD approach has been deemed a scientifically more advanced method for deriving a PoD however, compared to the NOAEL approach by several authorities. For example, the European Food Safety Authority (EFSA) has published an updated guidance on the use of the benchmark dose approach in risk assessment (EFSA, 2022), which gives guidance on how to apply the principles. Considering this, and for designing in vivo studies to qualify impurities, it is recommended to include at least four treated dose groups (besides a vehicle group) as well as 3 analysable animals/sex/group to ensure sufficient study power for modelling the dose-response data from the experimental animal studies. The principles of the EFSA BMD guidance on dose selection (EFSA, 2022) are recommended to ensure sufficient statistical relevance of the dose-response curve for deriving a BMDL. To qualify NMI, preliminary dose-range finding studies with the impurity are not recommended. Considering the usual AFs for deriving an AL, the BMDL used as PoD should be at least 500-fold higher than the anticipated AL using the clinical route of administration. When a different route of administration is used in the toxicology study, an appropriate multiple should be considered, taking into account the need to use A6 as an additional AF to account for differences in bioavailability. Generally, 3-fold increments are acceptable for spacing of the doses. The high dose may also be limited by the maximum tolerated dose.

Table 4: Preferred design of in vivo studies for qualification of impurities

Parameter	**Description**
Test substance	Neat (isolated impurity without API), purity > 95%.
Study design	GLP compliant and adhere to principles of OECD guideline 407
Duration of study/administration route	28-days (14-days for short term administration) and no recovery period. Administered via clinical route of administration.
Species/sex	Rats, unless otherwise justified. Both sexes should be included unless the clinical use of the medicinal product is only in 1.
Animals per group/number of groups	3 rats/sex/group. 4 dose groups. The highest dose level should be established with a suitable exposure margin compared to the proposed specification level, with the second highest dose group projected at the anticipated specification level, multiplied by the relevant AL-related AFs.
Control groups	Vehicle control group
TK analysis	3 M/F should be included for TK analysis. The analysis can be integrated in the main study as part of the high dose group.
Dose range	The dose range should ensure that the potential BMDL used as PoD is at least 500-fold higher than the anticipated AL

Special considerations for oncology products

For products covered under ICH S9, the general approach to qualify NMIs as described above may not be applicable. In line with ICH S9, impurity qualification may instead be based on the similarity in safety profile between the impurity and the DS. This approach typically results in a low level of concern

for most impurities in ICH S9 products (see Section 5.4.), and additional assessment is only warranted for impurities with a high level of concern; there is no need to control the impurity at a level where no toxicity is anticipated.

5.7.2. Setting limits based on in vivo data

Currently, no dedicated guidance has been issued by EMA or ICH regarding the derivation of a BMDL for use as PoD in establishing an AL for impurities. Until such guidance becomes available, the EFSA guidance from 2022 may be consulted for key principles and methodological considerations. Based on the best fitted model and the most relevant BMDL, an AL can be derived following the approach outlined in Section 5.5. of this reflection paper.

5.8. Products under clinical development

Evaluation of NMIs in investigational medicinal products under clinical development should follow the principles outlined in ICH M3(R2). The risk assessment may be supported by relevant sections of this reflection paper, as outlined above. Special consideration should be given to impurities of higher concern, such as those with known structural alerts or based on prior knowledge of related chemical classes. These may necessitate additional qualification efforts, such as lower batch-level usage or inclusion of supporting toxicological data. The level of concern analysis, as described in Section 5.4., can be a useful tool to determine the extent of qualification needed. Notably, short-term exposure in early-phase clinical trials may serve as a de-risking factor, potentially reducing the need for extensive impurity qualification. If the levels of NMIs in clinical trial batches are assessed – using the level of concern analysis - to be of low concern no further qualification efforts would be expected.

6. Conclusion

When impurity-specific safety information for an NMI may be necessary, alternative strategies beyond the conduct of additional in vivo general toxicology studies to gather this information may be followed. These alternative strategies include the use of TTC, (Q)SAR, read-across and in vitro approaches. This information can be used in an integrated risk assessment. A weight-of-evidence approach that includes all aspects that determine the level of concern, could be sufficient to decide that the NMI can be considered safe at the specified level.

7. References

Ajisafe, O.M., Adekunle, Y.A., Egbon, E., Ogbonna, C.E., Olawade, D.B., 'The role of machine learning in predictive toxicology: A review of current trends and future perspectives.', Life Sciences, 2025, 378, 123821, https://doi.org/10.1016/j.lfs.2025.123821.

Arnot, J.A., Toose, L., Armitage, J.M., Sangion, A., Looky, A., Brown, T.N., Li, L., Becker, R.A., 'Developing an internal threshold of toxicological concern (iTTC).', J. Expo Sci. Environ. Epidemiol., 2022, 32(6), 877-884, https://doi.org/10.1038/s41370-022-00494-x.

Ball, D., Blanchard, J., Jacobson-Kram, D., McClellan, R.O., McGovern, T., Norwood, D.L., Vogel, W., Wolff, R., Nagao, L., 'Development of safety qualification thresholds and their use in orally inhaled and nasal drug product evaluation.', Toxicol. Sci., 2007, 97(2), 226-236. https://doi.org/10.1093/toxsci/kfm058.

Bassan, A., Alves, V.M., Amberg, A., Anger, L.T., Beilke, L., Bender, A., Bernal, A., Cronin, M.T.D., Hsieh, J.H., Johnson, C., Kemper, R., Mumtaz, M., Neilson, L., Pavan, M., Pointon, A., Pletz, J., Ruiz, P., Russo, D.P., Sabnis, Y., Sandhu, R., Schaefer, M., Stavitskaya, L., Szabo, D.T., Valentin, J.P., Woolley, D., Zwickl, C., Myatt, G.J., 'In silico approaches in organ toxicity hazard assessment: Current status and future needs for predicting heart, kidney and lung toxicities.', Comput. Toxicol., 2021a, 20, 100188. https://doi.org/10.1016/j.comtox.2021.100188

Bassan, A., Alves, V.M., Amberg, A., Anger, L.T., Auerbach, S., Beilke, L., Bender, A., Cronin, M.T.D., Cross, K.P., Hsieh, J.H., Greene, N., Kemper, R., Kim, M.T., Mumtaz, M., Noeske, T., Pavan, M., Pletz, J., Russo, D.P., Sabnis, Y., Schaefer, M., Szabo, D.T., Valentin, J.P., Wichard, J., Williams, D., Woolley, D., Zwickl, C., Myatt, G.J., 'In silico approaches in organ toxicity hazard assessment: current status and future needs in predicting liver toxicity.', Comput. Toxicol., 2021b, 20, 100187. https://doi.org/10.1016/j.comtox.2021.100187

Bender, A., Scheiber, J., Glick, M., Davies, J.W., Azzaoui, K., Hamon, J., Urban, L., Whitebread, S., Jenkins, J.L., 'Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure.', Chem. Med. Chem., 2007, 2(6), 861-873. https://doi.org/10.1002/cmdc.200700026

Blum, J., Masjosthusmann, S., Bartmann, K., Bendt, F., Dolde, X., Dönmez, A., Förster, N., Holzer, A.K., Hübenthal, U., Keßel, H.E., Kilic, S., Klose, J., Pahl, M., Stürzl, L.C., Mangas, I., Terron, A., Crofton, K.M., Scholze, M., Mosig, A., Leist, M., Fritsche, E., 'Establishment of a human cell-based in vitro battery to assess developmental neurotoxicity hazard of chemicals.' Chemosphere, 2023, 311(Pt 2), 137035. https://doi.org/10.1016/j.chemosphere.2022.137035

Bowes, J., Brown, A.J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., Whitebread, S., 'Reducing safety-related drug attrition: the use of in vitro pharmacological profiling.', Nat. Rev. Drug Discov., 2012, 11(12), 909-922. https://doi.org/10.1038/nrd3845

Brennan, R.J., Jenkinson, S., Brown, A., Delaunois, A., Dumotier, B., Pannirselvam, M., Rao, M., Ribeiro, L.R., Schmidt, F., Sibony, A., Timsit, Y., Sales, V.T., Armstrong, D., Lagrutta, A., Mittlestadt, S.W., Naven, R., Peri, R., Roberts, S., Vergis, J.M., Valentin, J.P., 'The state of the art in secondary pharmacology and its impact on the safety of new medicines.', Nat. Rev. Drug Discov., 2024, 23(7), 525-545. https://doi.org/10.1038/s41573-024-00942-3

Cayley, A.N., Foster, R.S., Brigo, A., Muster, W., Musso, A., Kenyon, M.O., Parris, P., White, A.T., Cohen-Ohana, M., Nudelman, R., Glowienke, S., 'Assessing the utility of common arguments used in expert review of in silico predictions as part of ICH M7 assessments.', Regul. Toxicol. Pharmacol., 2023, 144, 105490. https://doi.org/10.1016/j.yrtph.2023.105490

Chilton, M.L., Api, A.M., Foster, R.S., Gerberick, G.F., Lavelle, M., Macmillan, D.S., Na, M., O'Brien, D., O'Leary-Steele, C., Patel, M., Ponting, D.J., Roberts, D.W., Safford, R.J., Tennant, R.E., 'Updating the Dermal Sensitisation Thresholds using an expanded dataset and an in silico expert system.', Regul. Toxicol. Pharmacol., 2022, 133, 105200. https://doi.org/10.1016/j.yrtph.2022.105200

Dik, S., Ezendam, J., Cunningham, A.R., Carrasquer, C.A., van Loveren, H., Rorije, E., 'Evaluation of in silico models for the identification of respiratory sensitizers.', Toxicol. Sci., 2014, 142(2), 385-394. https://doi.org/10.1093/toxsci/kfu188

EFSA Scientific Committee, 'Guidance on the use of the Threshold of Toxicological Concern approach in food safety assessment.', EFSA Journal, 2019, 17(6), 5708. https://doi.org/10.2903/j.efsa.2019.5708

EFSA Scientific Committee. 'Guidance on the use of the benchmark dose approach in risk assessment.' EFSA journal, 2022, 20(10), 7584. https://doi.org/10.2903/j.efsa.2022.7584

Eriksson, T., Björkman, S., Roth, B., Höglund, P., 'Intravenous formulations of the enantiomers of thalidomide: pharmacokinetic and initial pharmacodynamic characterization in man.', J. Pharm. Pharmacol., 2000, 52(7), 807-817. https://doi.org/10.1211/0022357001774660

Escher, S.E., Tluczkiewicz, I., Batke, M., Bitsch, A., Melber, C., Kroese, E.D., Buist, H.E., Mangelsdorf, I., 'Evaluation of inhalation TTC values with the database RepDose.', Regul. Toxicol. Pharmacol., 2010, 58(2), 259-274. https://doi.org/10.1016/j.yrtph.2010.06.009

Escher, S.E., Aguayo-Orozco, A., Benfenati, E., Bitsch, A., Braunbeck, T., Brotzmann, K., Bois, F., van der Burg, B., Castel, J., Exner, T., Gadaleta, D., Gardner, I., Goldmann, D., Hatley, O., Golbamaki, N., Graepel, R., Jennings, P., Limonciel, A., Long, A., Maclennan, R., Mombelli, E., Norinder, U., Jain, S., Capinha, L.S., Taboureau, O.T., Tolosa, L., Vrijenhoek, N.G., van Vugt-Lussenburg, B.M.A., Walker, P., van de Water, B., Wehr, M., White, A., Zdrazil, B., Fisher, C., 'Integrate mechanistic evidence from new approach methodologies (NAMs) into a read-across assessment to characterise trends in shared mode of action.'. Toxicol. In vitro., 2022, 79, 105269. https://doi.org/10.1016/j.tiv.2021.105269

Graham, J.C., Powley, M.W., Udovic, E., Glowienke, S., Nicolette, J., Parris, P., Kenyon, M., White, A., Maisey, A., Harvey, J., Martin, E.A., Dowdy, E., Masuda-Herrera, M., Trejo-Martin, A., Bercu J., 'Calculating qualified non-mutagenic impurity levels: Harmonization of approaches.', Regul. Toxicol. Pharmacol., 2021, 126, 105023. https://doi.org/10.1016/j.yrtph.2021.105023

Golden, E., Macmillan, D.S., Dameron, G., Kern, P., Hartung, T., Maertens, A., 'Evaluation of the global performance of eight in silico skin sensitization models using human data.' ALTEX, 2021, 38(1), 33-48. https://doi.org/10.14573/altex.1911261

Harvey, J., Fleetwood, A., Ogilvie, R., Teasdale, A., Wilcox, P., Spanhaak, S., 'Management of organic impurities in small molecule medicinal products: Deriving safe limits for use in early development.', Regul. Toxicol. Pharmacol., 2017, 84, 116-123. https://doi.org/10.1016/j.yrtph.2016.12.011

Hasselgren, C., Kenyon, M., Anger, L.T., Cornwell, P., Watt, E., Bercu, J., 'Analysis of non-mutagenic substances in the context of drug impurity assessment – few are potent toxicants.', Regul. Toxicol. Pharmacol., 2024, 150, 105645. https://doi.org/10.1016/j.yrtph.2024.105645

Hasselgren, C., Ahlberg, E., Akahori, Y., Amberg, A., Anger, L.T., Atienzar, F., Auerbach, S., Beilke, L., Bellion, P., Benigni, R., Bercu, J., Booth, E.D., Bower, D., Brigo, A., Cammerer, Z., Cronin, M.T.D., Crooks, I., Cross, K.P., Custer, L., Dobo, K., Doktorova, T., Faulkner, D., Ford, K.A., Fortin, M.C., Frericks, M., Gad-McDonald, S.E., Gellatly, N., Gerets, H., Gervais, V., Glowienke, S., Van Gompel, J., Harvey, J.S., Hillegass, J., Honma, M., Hsieh, J.H., Hsu, C.W., Barton-Maclaren, T.S., Johnson, C., Jolly, R., Jones, D., Kemper, R., Kenyon, M.O., Kruhlak, N.L., Kulkarni, S.A., Kümmerer, K., Leavitt, P., Masten, S., Miller, S., Moudgal, C., Muster, W., Paulino, A., Lo Piparo, E., Powley, M., Quigley, D.P.,

Reddy, M.V., Richarz, A.N., Schilter, B., Snyder, R.D., Stavitskaya, L., Stidl, R., Szabo, D.T., Teasdale, A., Tice, R.R., Trejo-Martin, A., Vuorinen, A., Wall, B.A., Watts, P., White, A.T., Wichard, J., Witt, K.L., Woolley, A., Woolley, D., Zwickl, C., Myatt, G.J., 'Genetic toxicology in silico protocol.', Regul. Toxicol. Pharmacol., 2019, 107, 104403. https://doi.org/10.1016/j.yrtph.2019.104403

Kenyon, M.O., Martin, M., Martin, E.A., Brandstetter, S., Wegesser, T., Greene, N., Harvey, J., 'Deriving acceptable limits for process-related organic impurities in medicinal products – Durational adjustments.', Regul. Toxicol. Pharmacol., 2024, 150, 105644. https://doi.org/10.1016/j.yrtph.2024.105644

Lynch, J.J. 3rd, Van Vleet, T.R., Mittelstadt, S.W., Blomme, E.A.G., 'Potential functional and pathological side effects related to off-target pharmacological activity.', J. Pharmacol. Toxicol. Methods, 2017, 87, 108-126. https://doi.org/10.1016/j.vascn.2017.02.020

Masuda-Herrera, M.J., Bercu, J.P., Broschard, T.H., Burild, A., Hasselgren, C., Parris, P., Ford, L.C., Graham, J., Stanard, B., Comerford, M., Lettiere, D., Erler, S., Callis, C.M., Morinello, E., Muster, W., Martin, E.A., Griffin, T.R., Nagao, L., Cruz, M., 'Development of Duration-Based Non-Mutagenic Thresholds of Toxicological Concern (TTCs) Relevant to Parenteral Extractables and Leachables (E&Ls).', PDA J. Pharm. Sci. Technol., 2022, 76(5), 369–383. https://doi.org/10.5731/pdajpst.2021.012693

Mitra, M.S., Datta, K., Hutchinson, R., Nicolette, J.J., Pettersen, J.C., Wegesser, T.C., Bercu, J.P., 'Harmonized 3Rs-based non-mutagenic impurity qualification study designs developed using the results of an IQ consortium survey.', Regul. Toxicol. Pharmacol., 2021, 122, 104895. https://doi.org/10.1016/j.yrtph.2021.104895

Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G., 'Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of toxicological concern.', Food Chem. Toxicol., 1996, 34, 829–867. https://doi.org/10.1016/S0278-6915(96)00049-X

Myatt, G.J., Ahlberg, E., Akahori, Y., Allen, D., Amberg, A., Anger, L.T., Aptula, A., Auerbach, S., Beilke, L., Bellion, P., Benigni, R., Bercu, J., Booth, E.D., Bower, D., Brigo, A., Burden, N., Cammerer, Z., Cronin, M.T.D., Cross, K.P., Custer, L., Dettwiler, M., Dobo, K., Ford, K.A., Fortin, M.C., Gad-McDonald, S.E., Gellatly, N., Gervais, V., Glover, K.P., Glowienke, S., Van Gompel, J., Gutsell, S., Hardy, B., Harvey, J.S., Hillegass, J., Honma, M., Hsieh, J.H., Hsu, C.W., Hughes, K., Johnson, C., Jolly, R., Jones, D., Kemper, R., Kenyon, M.O., Kim, M.T., Kruhlak, N.L., Kulkarni, S.A., Kümmerer, K., Leavitt, P., Majer, B., Masten, S., Miller, S., Moser, J., Mumtaz, M., Muster, W., Neilson, L., Oprea, T.I., Patlewicz, G., Paulino, A., Lo Piparo, E., Powley, M., Quigley, D.P., Reddy, M.V., Richarz, A.N., Ruiz, P., Schilter, B., Serafimova, R., Simpson, W., Stavitskaya, L., Stidl, R., Suarez-Rodriguez, D., Szabo, D.T., Teasdale, A., Trejo-Martin, A., Valentin, J.P., Vuorinen, A., Wall, B.A., Watts, P., White, A.T., Wichard, J., Witt, K.L., Woolley, A., Woolley, D., Zwickl, C., Hasselgren, C., 'In silico toxicology protocols.', Regul. Toxicol. Pharmacol., 2018, 96, 1–17. https://doi.org/10.1016/j.yrtph.2018.04.014

Myatt, G.J., Beilke, L.D., Cross, K.P., 'In silico tools and their application.', Comp. Med. Chem. III, 2017, 156-176. https://doi.org/10.1016/B978-0-12-409547-2.12379-0

Nelms, M.D., Patlewicz, G., 'Derivation of New Threshold of Toxicological Concern Values for Exposure via Inhalation for Environmentally-Relevant Chemicals.', Front. Toxicol., 2020, 2, 580347. https://doi.org/10.3389/ftox.2020.580347

Nishijo, T., Api, A.M., Gerberick, G.F., Miyazawa, M., Roberts, D.W., Safford, R.J., Sakaguchi, H., 'Application of the dermal sensitization threshold concept to chemicals classified as high potency category for skin sensitization assessment of ingredients for consumer products.', Regul. Toxicol. Pharmacol., 2020, 117, 104732. https://doi.org/10.1016/j.yrtph.2020.104732

OECD (2004), 'OECD principles for the Validation, for Regulatory Purposes, of (Q)SAR Models', OECD Series on Testing and Assessment, No, 49, OECD Publishing, Paris, https://one.oecd.org/document/ENV/JM/MONO(2004)24/en/pdf

OECD (2014), 'Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models', OECD Series on Testing and Assessment, No. 69, OECD Publishing, Paris, https://doi.org/10.1787/9789264085442-en

OECD (2017), *'Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA)',* OECD Series on Testing and Assessment, No. 260, OECD Publishing, Paris, https://doi.org/10.1787/44bb06c1-en

OECD (2023), `(Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions', OECD Series on Testing and Assessment, No. 386, OECD Publishing, Paris, https://doi.org/10.1787/d96118f6-en

OECD (2024), '(Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions, Second Edition', OECD Series on Testing and Assessment, No. 405, OECD Publishing, Paris, https://doi.org/10.1787/bbdac345-en

Paixão, P., Gouveia, L.F., Morais, J.A., 'Prediction of the human oral bioavailability by using in vitro and in silico drug related parameters in a physiologically based absorption model.', Int. J. Pharm., 2012, 429(1–2), 84–98. https://doi.org/10.1016/j.ijpharm.2012.03.019

Parris, P., Whelan, G., Burild, A., Whritenour, J., Bruen, U., Bercu, J., Callis, C., Chilton, M.L., Graham, J., Johann, E., Johnson, C., Griffin, T., Kohan, M., Martin, E.A., Masuda-Herrera, M., Stanard, B., Cruz, M.T., Nagao, L., 'Sensitization Assessment of Extractables and Leachables in Pharmaceuticals: ELSIE Database Analysis.', PDA J. Pharm. Sci. Technol., 2024, 78(4), 399-444. https://doi.org/10.5731/pdajpst.2022.012811

Partosch, F., Mielke, H., Stahlmann, R., Kleuser, B., Barlow, S., Gundert-Remy, U., 'Internal threshold of toxicological concern values: enabling route-to-route extrapolation.', Arch. Toxicol., 2015, 89(6), 941–948. https://doi.org/10.1007/s00204-014-1287-6

Raies, A.B. and Bajic, V.B., 'In silico toxicology: computational methods for the prediction of chemical toxicity.', WIREs Comput. Mol. Sci., 2016, 6, 147–172. https://doi.org/10.1002/wcms.1240

Rim, K.T., 'In silico prediction of toxicity and its applications for chemicals at work.', Toxicol. Environ. Health Sci., 2020, 12, 191–202. https://doi.org/10.1007/s13530-020-00056-4

Roberts, D.W., Api, A.M., Safford, R.J., Lalko, J.F., 'Principles for identification of High Potency Category Chemicals for which the Dermal Sensitisation Threshold (DST) approach should not be applied.', Regul. Toxicol. Pharmacol., 2015, 72(3), 683–693. https://doi.org/10.1016/j.yrtph.2015.03.001

Rovida, C., Escher, S.E., Herzler, M., Bennekou, S.H., Kamp, H., Kroese, D.E., Maslankiewicz, L., Moné, M.J., Patlewicz, G., Sipes, N., Van Aerts, L., White, A., Yamada, T., Van de Water, B., 'NAM-supported read-across: From case studies to regulatory guidance in safety assessment.', ALTEX, 2021, 38(1), 140–150. https://doi.org/10.14573/altex.2010062

Safford, R.J., 'The Dermal Sensitisation Threshold – a TTC approach for allergic contact dermatitis.', Regul. Toxicol. Pharmacol., 2008, 51(2), 195–200. https://doi.org/10.1016/j.yrtph.2008.02.010

Safford, R.J., Api, A.M., Roberts, D.W., Lalko, J.F., 'Extension of the Dermal Sensitisation Threshold (DST) approach to incorporate chemicals classified as reactive.', Regul. Toxicol. Pharmacol., 2015, 72(3), 694–701. https://doi.org/10.1016/j.yrtph.2015.04.020

Schirris, T.J., Renkema, G.H., Ritschel, T., Voermans, N.C., Bilos, A., van Engelen, B.G., Brandt, U., Koopman, W.J., Beyrath, J.D., Rodenburg, R.J., Willems, P.H., Smeitink, J.A., Russel, F.G., 'Statin-Induced Myopathy Is Associated with Mitochondrial Complex III Inhibition.', Cell Metab., 2015, 22(3), 399–407. https://doi.org/10.1016/j.cmet.2015.08.002

Schmeisser, S., Miccoli, A., von Bergen, M., Berggren, E., Braeuning, A., Busch, W., Desaintes, C., Gourmelon, A., Grafström, R., Harrill, J., Hartung, T., Herzler, M., Kass, G.E.N., Kleinstreuer, N., Leist, M., Luijten, M., Marx-Stoelting, P., Poetz, O., van Ravenzwaay, B., Roggeband, R., Rogiers, V., Roth, A., Sanders, P., Thomas, R.S., Vinggaard, A.M., Vinken, M., van de Water, B., Luch, A., Tralau, T., 'New approach methodologies in human regulatory toxicology – Not if, but how and when!', Environ. Int., 2023, 178, 108082. https://doi.org/10.1016/j.envint.2023.108082

Schneckener, S., Grimbs, S., Hey, J., Menz, S., Osmers, M., Schaper, S., Hillisch, A., Göller, A.H., 'Prediction of Oral Bioavailability in Rats: Transferring Insights from in vitro Correlations to (Deep) Machine Learning Models Using in silico Model Outputs and Chemical Structure Parameters.', J Chem Inf Model., 2019, 59(11), 4893–4905. https://doi.org/10.1021/acs.jcim.9b00460

Sheridan, R.P. and Kearsley, S.K., 'Why do we need so many chemical similarity search methods?', Drug Discov. Today, 2002, 7(17), 903–911. https://doi.org/10.1016/S1359-6446(02)02412-1

Slikkerveer, A., Doehr, O., Claude, N., Hutchinson, R., Harvey, J., Spanhaak, S., 'New limits proposed for the management of non-mutagenic impurities.', Regul. Toxicol. Pharmacol., 2024, 150, 105647. https://doi.org/10.1016/j.yrtph.2024.105647

Stice, S. and Adams, T.B., 'The Expanded Decision Tree: Predicting Chronic Toxic Potential and Safe Intake Levels', Food Chem. Toxicol., 2025, 115903, https://doi.org/10.1016/j.fct.2025.115903

Stielow, M., Witczyńska, A., Kubryń, N., Fijałkowski, Ł., Nowaczyk, J., Nowaczyk, A., 'The Bioavailability of Drugs – The Current State of Knowledge.', Molecules, 2023, 28(24), 8038. https://doi.org/10.3390/molecules28248038

Stumpfe, D., Hu, H., Bajorath, J., 'Evolving Concept of Activity Cliffs.', ACS Omega., 2019, 4(11), 14360–14368. https://doi.org/10.1021/acsomega.9b02221

Ta, G.H., Weng, C.F., Leong, M.K., 'In silico Prediction of Skin Sensitization: Quo vadis?', Front Pharmacol., 2021, 12, 655771. https://doi.org/10.3389/fphar.2021.655771

Teubner, W., Mehling, A., Schuster, P.X., Guth, K., Worth, A., Burton, J., van Ravenzwaay, B., Landsiedel, R., 'Computer models versus reality: how well do in silico models currently predict the sensitization potential of a substance.', Regul. Toxicol. Pharmacol., 2013, 67(3), 468–485. https://doi.org/10.1016/j.yrtph.2013.09.007

Tice, R.R., Bassan, A., Amberg, A., Anger, L.T., Beal, M.A., Bellion, P., Benigni, R., Birmingham, J., Brigo, A., Bringezu, F., Ceriani, L., Crooks, I., Cross, K., Elespuru, R., Faulkner, D.M., Fortin, M.C., Fowler, P., Frericks, M., Gerets, H.H.J., Jahnke, G.D., Jones, D.R., Kruhlak, N.L., Lo Piparo, E., Lopez-Belmonte, J., Luniwal, A., Luu, A., Madia, F., Manganelli, S., Manickam, B., Mestres, J., Mihalchik-Burhans, A.L., Neilson, L., Pandiri, A., Pavan, M., Rider, C.V., Rooney, J.P., Trejo-Martin, A., Watanabe-Sailor, K.H., White, A.T., Woolley, D., Myatt, G.J., 'In Silico Approaches In Carcinogenicity Hazard Assessment: Current Status and Future Needs.', Comput. Toxicol., 2021, 20, 100191. https://doi.org/10.1016/j.comtox.2021.100191

Tluczkiewicz, I., Buist, H.E., Martin, M.T., Mangelsdorf, I., Escher, S.E., 'Improvement of the Cramer classification for oral exposure using the database TTC RepDose – A strategy description.', Regul. Toxicol. Pharmacol., 2011, 61, 340–350. https://doi.org/10.1016/j.yrtph.2011.09.005

Tluczkiewicz, I., Kühne, R., Ebert, R.U., Batke, M., Schüürmann, G., Mangelsdorf, I., Escher, S.E., 'Inhalation TTC values: A new integrative grouping approach considering structural, toxicological and mechanistic features.', Regul. Toxicol. Pharmacol., 2016, 78, 8–23. https://doi.org/10.1016/j.yrtph.2016.03.022

Tobias, A., Ballard, B.D., Mohiuddin, S.S., 'Physiology, Water Balance', In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2025. Available from: https://www.ncbi.nlm.nih.gov/books/NBK541059/

Wang, C.C., Lin, Y.C., Wang, S.S., Shih, C., Lin, Y.H., Tung, C.W., 'SkinSensDB: a curated database for skin sensitization assays', J. Cheminform., 2017, 9, 5. https://doi.org/10.1186/s13321-017-0194-2

Watford, S., Ly Pham, L., Wignall, J., Shin, R., Martin, M.T., Friedman, K.P., 'ToxRefDB version 2.0: Improved utility for predictive and retrospective toxicology analyses.', Reprod. Toxicol., 2019, 89, 145–158. https://doi.org/10.1016/j.reprotox.2019.07.012

Wei, Z., Xu, T., Strickland, J., Zhang, L., Fang, Y., Tao, D., Simeonov, A., Huang, R., Kleinstreuer, N.C., Xia, M., 'Use of in vitro methods combined with in silico analysis to identify potential skin sensitizers in the Tox21 10K compound library.', Front. Toxicol., 2024, 6, 1321857. https://doi.org/10.3389/ftox.2024.1321857

Weidolf, L., Andersson, T., Bercu, J.P., Brink, A., Glowienke, S., Harvey, J., Hayes, M.A., Jacques, P., Lu, C., Manevski, N., Muster, W., Nudelman, R., Ogilvie, R., Ottosson, J., Teasdale, A., Trela, B., 'Qualification of impurities based on metabolite data.', Regul. Toxicol. Pharmacol., 2020, 110, 104524. https://doi.org/10.1016/j.yrtph.2019.104524

Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., Richard, A.M., 'The CompTox Chemistry Dashboard: a community data resource for environmental chemistry.', J. Cheminform., 2017, 9(1), 61. https://doi.org/10.1186/s13321-017-0247-6

Willett, P., Barnard, J.M., Downs, G.M., 'Chemical similarity searching.', J. Chem. Inf. Comput. Sci., 1998, 38(6), 983–996. https://doi.org/10.1021/ci9800211.

Worth, A., Lapenna, S., Piparo, L.E., Mostrag-Szlichtyng, A., Serafimova, R., 'A framework for assessing in silico toxicity predictions: Case studies with selected pesticides.', EUR 24705 EN. Luxembourg (Luxembourg): Publications Office of the European Union, 2011, JRC62586. *https://dx.doi.org/10.2788/29048*

Xiao, K., 'Degradation Chemistry and Product Development', In Analytical Scientists in Pharmaceutical Product Development, K. Xiao (Ed.), 2020, https://doi.org/10.1002/9781119547785.ch5

Annex 1

Use of assessment factors

The use of modifying factors for the derivation of a PDE is described in ICH Q3C/D/E documents. Here, we reflect on the use of AFs, which are similar to the modifying factors described for the PDE methodology, but these reflections do not necessarily apply to the PDE methodology since for the derivation of an AL, product-specific considerations are taken into account.

AF1 is a factor to account for extrapolation between species.

The use of AF1 is the same as for F1 in the PDE methodology. F1 considers the comparative surface area:body weight ratios for the species concerned and for man. Surface area (S) is calculated as:

$$S = kM^{0.67}$$

in which M = body mass, and the constant k is set to 10. Standard values for common laboratory species are available in ICH guidelines. 50 kg is used as the standard body weight for humans. In some cases, a value is not provided in the ICH document, or there are reasons to deviate. In these cases, an AF1 can be calculated using the formula:

$$1/(M_{animal}/M_{human})^{0.33}$$

For instance, when developmental neurotoxicity is taken as a PoD from a rat juvenile study in which the average body weight of the pups was 14 g, this needs to be compared with the average body weight of infants, 5 kg (EFSA, 2012). Using the above formula provides an AF1 of 7.

AF2 is a factor of 10 to account for variability between individuals

A factor of 10 is generally applied (ICH Q3C/ICH Q3D). ICH Q3D provides the possibility to split this factor into a toxicokinetic and a toxicodynamic component (3.18 each) on a case-by-case basis. It is less likely that this option can be used to justify an AL for NMIs. It is difficult to justify that there is little or no toxicodynamic differences between individuals in the target population. Toxicokinetics of organic compounds is more complex than of elemental impurities and includes metabolism. Often it is not known if a potential toxicological hazard is attributable to the parent impurity or to a metabolite. Besides metabolism, variability in absorption, distribution, and elimination may exist within the target population.

AF3 is a variable factor to account for toxicity studies of short-term exposure

Specific values for F3 mentioned in Q3C/D, to be used when the PoD is taken from a non-chronic study, in principle are also applicable when deriving an AL for a specific product. However, some changes can be considered on a case-by-case basis, for instance taking into account that the product is administered intermittently, or only short-term (*i.e.* up to 1 month). In the case of medicinal products administered for less than a patient's lifetime, it may be appropriate to select a PoD from an animal study with a relatively short duration and use a lower value for AF3 than would usually be applied when a PDE is derived for chronic use. If additional animal studies are available with even longer duration, these may have BMDL/NOAEL values based on findings that may not be relevant for shorter term exposures and therefore may not be the most appropriate PoD for a given DP.

AF4 is a factor that may be applied in cases of severe toxicity, e.g., non-mutagenic carcinogenicity, neurotoxicity, or teratogenicity.

This factor can be used to take into account the severity of an effect. Some examples for developmental toxicity are provided in ICH Q3C/D. When no BMDL or NOAEL is available and a LOAEL is used at which the severe effect is observed, a high AF4 value up to 10 should be considered.

The value for AF4 is always related to the severity of the adverse effects at the LOAEL that is chosen as critical endpoint. This is illustrated in the example below.

Table 5: Example for choosing AF4 depending on severity and critical endpoint.

Dose level	**Adverse effects**	**Critical endpoint**	
		liver toxicity	**CNS toxicity**
1 mg/kg/day	None	NOAEL	
3 mg/kg/day	Significant changes in liver enzymes	LOAEL	NOAEL
10 mg/kg/day	Significant increased liver weight and convulsions		LOAEL

In this example, when choosing liver toxicity as a relevant endpoint, the PoD would be 1 mg/kg/day and AF4 would be 1. When in the same study CNS toxicity is observed and considered the critical endpoint, the PoD would be 3 mg/kg/day and AF4 would be 10. CNS toxicity would then provide the lower PDE and be considered the most relevant endpoint for deriving the PDE for this compound.

The absence of data is not a reason to set a value greater than 1 for AF4. For instance, when no developmental toxicity data are available, AF4 should not be assigned a value of 10 just because the absence of data cannot exclude the possibility of a teratogenic effect. Absence and quality of data is something that needs to be considered as part of an uncertainty analysis in any risk assessment.

AF5 is a variable factor that may be applied if the NOEL was not established

ICH Q3C indicates that a factor of up to 10 could be used depending on the severity of the toxicity if the no-effect level (NOEL) was not established. ICH Q3D differentiates between NOAEL/LOAEL and NOEL/lowest observed effect level (LOEL). Both guidelines indicate that the severity/adversity of the effects plays a role in determining AF5. Yet, as discussed above, AF4 also addresses the severity of the effects observed. For AF5, the steepness of the dose-response curve is also relevant in choosing the values. In cases of a shallow dose-response curve between the NOAEL and the LOAEL a higher AF5 of up to 5 may be appropriate. This will however also depend on the spacing between doses, where a close spacing between the NOAEL and LOAEL will lead to a high AF5, e.g. 5, and a wider dose spacing will lead to a lower AF5, e.g. 3-4. In cases of a steep dose-response curve between NOAEL and LOAEL, AF5 may be more moderate. Here, a close dose spacing could lead to a higher AF5, e.g. 2-3, while a wide dose spacing between NOAEL and LOAEL would lead to a low AF5, e.g. 1.

If an effect is observed at the PoD but is not considered adverse, this dose level is considered to be a NOAEL, but it also can be considered to be a LOEL. This should not lead to a value greater than 1 for AF5. The observed effect could be an adaptive response without any adverse sequelae. However, when the effect at the LOEL is related to the same process that leads to adversity at higher dose levels or with longer duration of exposure, it could be justifiable to assign an AF5 level greater than 1 (e.g. 3). If sufficient information is available, the choice of AF5 may also be informed by the distance between the LOAEL and the projected dose level where no adverse effects are seen. In case a BMDL is chosen as PoD and the BMDL is sufficiently justified, the value for AF5 would be 1.

Based on these considerations, the choice of values for AF4 and AF5 are related to some extent. Choosing a value of 10 for both factors would only be appropriate when severe toxicity is observed at the LOAEL, and this dose level is used as PoD.

AF6 is a variable factor to account for route of exposure difference (e.g., oral versus parenteral).

In the absence of data for the intended route of exposure and/or where data are available but not considered sufficient for a safety assessment for the route of administration, a modifying factor can be used to correct for a difference in bioavailability between the route used in the study from which the PoD is taken and the bioavailability for the route for which an AL is being derived. Ideally, AF6 should be based on bioavailability of the parent compound. If a radiolabel study is used, it should be referred to as absorption because it is not clear if the radiolabel is the parent, or a metabolite, or a combination of parent and metabolite(s). For example, when adequate data suggests that the oral bioavailability is 30% and the PoD is taken from an oral toxicity study whereas the product is administered parenterally, AF6 would be 3. Alternatively, default factors can be applied when the bioavailability estimate is uncertain (due to conflicting data, the use of data with limited reliability, or dependence on bioavailability data for a surrogate compound). When using oral toxicity data to derive a parenteral AL conversion factor similar to Q3D can be used:

AF6= 100 Oral bioavailability <1%: divide by a modifying factor of 100;

AF6= 10 Oral bioavailability ≥ 1% and <50%: divide by a modifying factor of 10;

AF6= 2 Oral bioavailability ≥50% and <90%: divide by a modifying factor of 2; and

AF6=1 Oral bioavailability ≥ 90%: divide by a modifying factor of 1.

In the absence of in vivo data, a NAM approach – combining in vitro data estimating oral absorption and internal clearance, with an in silico PBPK model - can be used to generate data for assessing bioavailability. The reliability of such models should be documented. When the compound is out of the applicability domain of the model, or when the reliability index is too low, the result of the model should be discarded. When sufficiently justified, the results from a NAM approach in regulatory submissions can be considered by the authorities.

Where appropriate bioavailability data were not available, and in lieu of NAM-derived estimates of bioavailability, a default modifying factor of 100 is suggested for AF6. Smaller values need further justification, e.g. reasoning based on the physicochemical characteristics of the compound. In addition, evidence of a clear biological response after oral exposure in toxicity studies can be leveraged to support a smaller AF6. When suitable bioavailability data are available for a surrogate molecule, allowing a read-across approach, these data may be leveraged to inform the bioavailability estimate, if sufficiently justified.

When the data concerns an inhalation toxicology study, data on respiratory tract deposition, respiratory absorption rate and pulmonary metabolism may inform on AF6. If such data are not available and a parenteral AL needs to be derived, the value for AF6 needs justification, e.g. based on physico-chemical properties. If a compound shows local toxicity in the absence of systemic toxicity, the dose at which these effects are observed is less suitable to derive a parenteral AL. In contrast, when the medicinal product is administered by inhalation and no inhalation toxicology data are available for the leachable, as a cautious approach, 100% bioavailability of the external dose can be assumed, and the inhalation AL would be the same as the parenteral AL. When data can be presented that show bioavailability is less, this could justify a smaller AF6.

Likewise, when systemic toxicity data observed in a dermal toxicity study are used to derive a parenteral AL and data on absorption are available after dermal exposure, AF6 can be based on these

absorption data. In the absence of actual absorption data, AF6 needs to be justified, e.g. based on physiochemical characteristics of the compound and the formulation.

AF7 is a variable factor that may be applied if a read-across strategy is used.

When read-across strategy is utilised, a factor of up to 5 could be used depending on the level of (dis)similarity. In general, when a surrogate is considered similar based on the criteria described in this guideline, an AF7 of 1 may be applicable.

Annex 2

Concept of a Risk Exposure Matrix

Step 1 and 2: Define Parameters and Their Risk Level

We use the key factors from Figure 2 but assign them an Inherent Risk Level.

Parameter	Tier	Rationale
1. Daily Intake	< 1.5 µg/day	Negligible exposure
(Exposure)	1.5 µg - 100 µg/day	Low exposure
	100 µg - 1 mg/day	Moderate exposure
	> 1 mg/day	Significant exposure
---	---	---
2. Duration of Treatment	< 1 month (acute)	Limited duration
	1 - 12 months (sub-chronic)	Intermediate duration
	> 12 months (chronic)	Long-term duration
---	---	---
3. Target Population	Adults (>18 years)	Standard population
	Elderly / Moderate impairment	Increased vulnerability
	Children / Severe impairment	High vulnerability
	Neonates / Pregnant women	Critical vulnerability

Step 3: Introduce a Modulating Factor (Benefit/Risk)

The clinical indication is not a risk itself but a **modulator of risk acceptability**. It may therefore affect the risk level.

4. Clinical Indication	Rationale
Life-threatening disease / No alternative	High tolerance for risk
Serious, non-life-threatening disease	Moderate tolerance
Non-serious / symptomatic condition	Low tolerance for risk

Step 4: Define Action Level

The determined Inherent Risk Level is then modulated by the clinical indication to define the final Level of Concern and its corresponding required action.

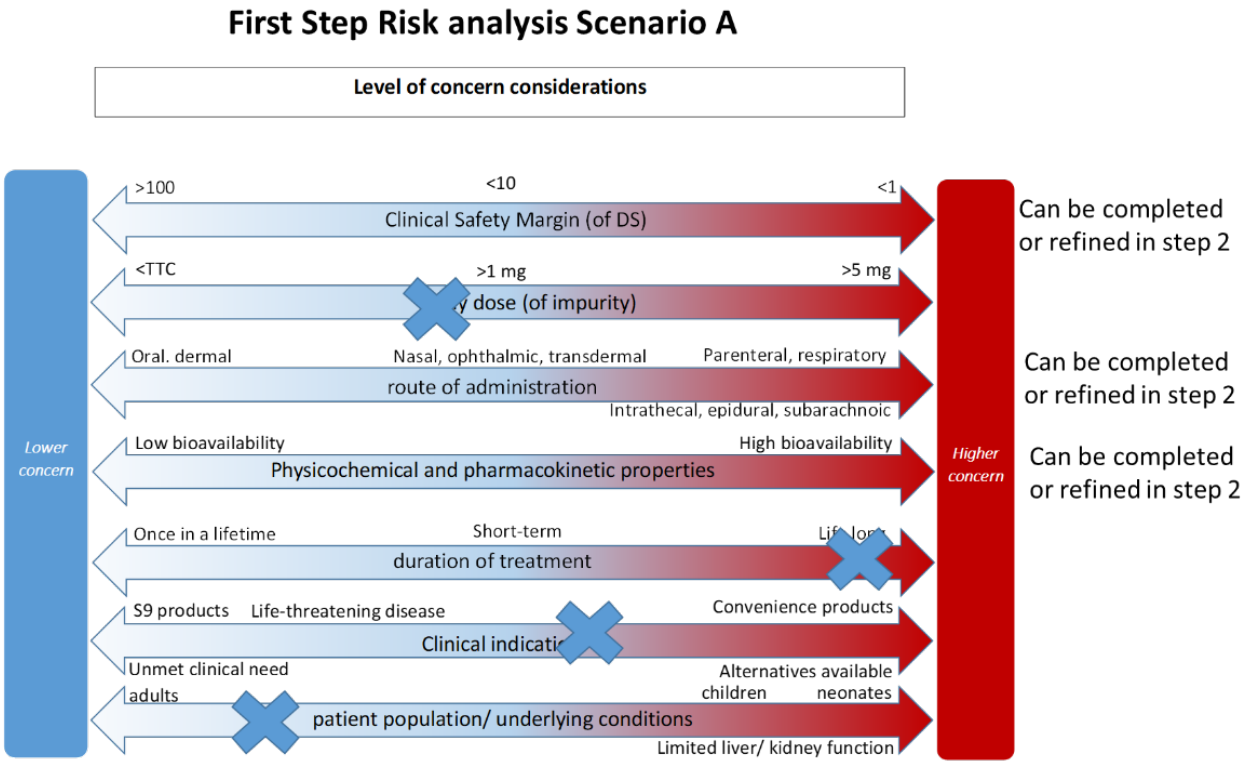
Level of Concern	Required Action
Low Concern	Qualification by simple justification.
Moderate Concern	In-depth assessment required (literature, in silico).
High Concern	Experimental toxicological data likely required.

Practical Examples of First Level of Concern analysis

Two scenarios are described below that would be difficult to differentiate using a purely qualitative approach.

Scenario A: Impurity in a chronic treatment for hypertension

- **Intake:** 150 µg/day
- **Duration:** Lifelong treatment (> 12 months)
- **Population:** Adults and elderly
- **Indication:** Serious, non-life-threatening disease



Step-by-Step Risk Assessment Through the Decision Flowchart

Step 1: **Initial Exposure Level determination**

- The intake of 150 µg/day is above the TTC threshold of 1.5 µg/day and falls into the *Moderate Exposure category.*

Step 2: **Determine Inherent Risk**

- Exposure: Moderate.
- Duration: Lifelong treatment (> 12 months) is a long-term exposure, which is a high-risk factor.
- Population: Adults and elderly (standard/increased vulnerability).
- The presence of a single high-risk factor (the long-term duration) is sufficient to classify *the overall risk as High Inherent Risk.*

Step 3: Modulation by Clinical Indication

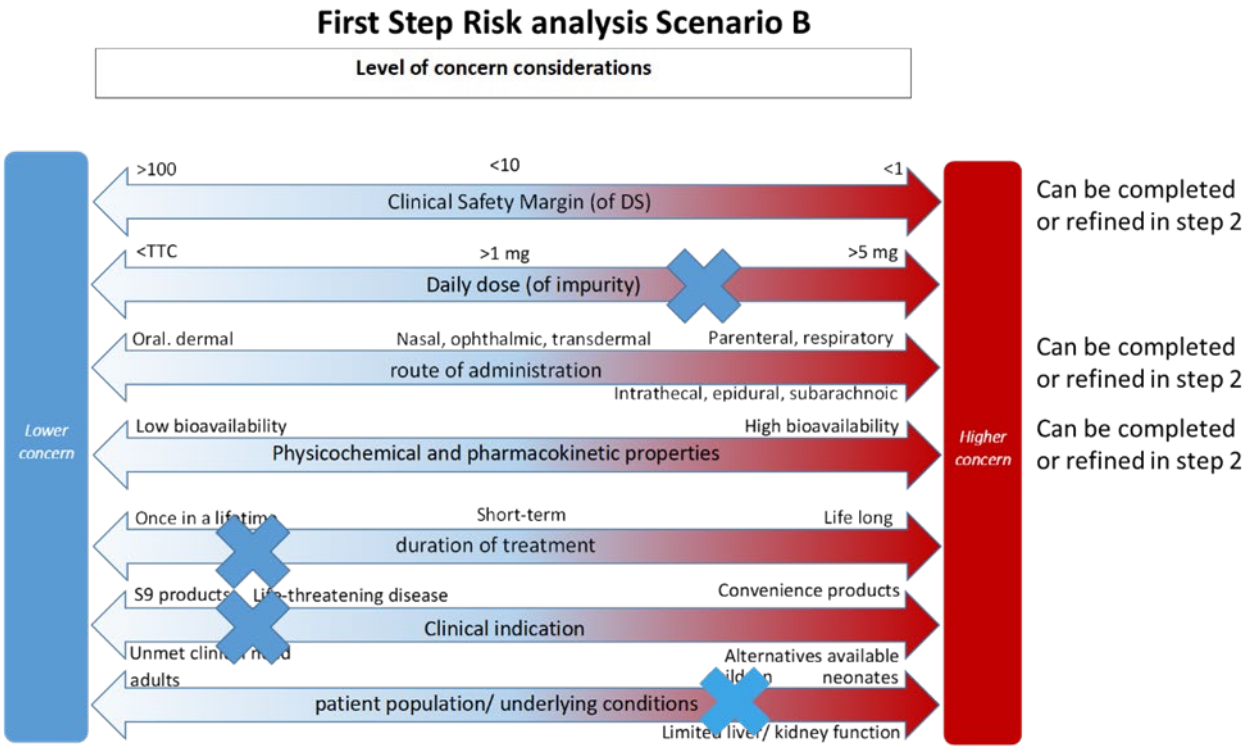
- The High Inherent Risk is now modulated by the indication: "Serious, non-life-threatening disease." The clinical benefit is significant, but not as critical as in a life-or-death situation.
- The decision tree therefore balances the high risk against this substantial benefit, which reduces the final level of concern to ***Moderate Concern***.

Step 4: Define Action Level

Conclusion: Moderate Concern: The outcome aligns with the analysis. It requires an in-depth assessment (e.g., in silico) but not necessarily immediate experimental data.

Scenario B: Impurity in a short-term chemotherapy for children

- **Intake:** 1.2 mg/day
- **Duration:** 3 weeks (< 1 month)
- **Population:** Children
- **Indication:** Life-threatening disease



Step-by-Step Risk Assessment Through the Decision Flowchart

Step 1: Initial Exposure Level determination

- The intake of 1.2 mg/day is in the highest category, Significant Exposure.

Step 2: Determine Inherent Risk

- Exposure: Significant.

- Duration: 3 weeks (< 1 month) is a limited duration.
- Population: Children are a high-vulnerability population.
- The presence of multiple high-risk factors (significant exposure and a vulnerable population) unequivocally leads to a High Inherent Risk.

Step 3: Modulation by Clinical Indication

- The High Inherent Risk is modulated by the indication: "Life-threatening disease."
- Here, the benefit/risk factor is at its maximum. The treatment is essential for survival. Therefore, even a high inherent risk is tolerated, which drastically reduces the final level of concern to Moderate Concern.

Moderate Concern. Despite a high dose in a sensitive population, the clinical context (benefit/risk) and the short duration drastically reduce the overall concern.

Step 4: Define Action Level

Conclusion: This scenario illustrates the power of the modulating factor. An intrinsically high toxicological risk becomes acceptable when the clinical benefit is critical, justifying the conclusion of Moderate Concern.