EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

1 02 December 2024
2 EMA/CHMP/543397/2024
3 Committee for Medicinal Products for Human Use (CHMP)

4 Reflection paper on the qualification of non-mutagenic 5 impurities

6 Draft

Draft agreed by NcWP	09 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

7

Comments should be provided using this EUSurvey form. For any technical issues, please contact the EUSurvey Support.

8

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	Adsorption, distribution, metabolism and excretion
AF	Assessment factor
AI	Artificial intelligence
AL	Acceptable level
AOP	Adverse outcome pathways
API	Active pharmaceutical ingredient
BMD	Benchmark dose
BMDL	Benchmark dose lower boundary
BMDU	Benchmark dose upper boundary
bw	Bodyweight
CHMP	Committee for Medicinal Products for Human use
CNS	Central nervous system
CVMP	Committee for Veterinary Medicinal Products
CVS	Cardiovascular system
DNA	Desoxyribonucleic acid
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
EWG	Expert working group
FDA	Food and Drug Administration
GIT	Gastrointestinal tract
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
MeOC	Median observed concentration
ML	Machine learning
MoA	Mode of action
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
PC	Physicochemical
PDE	Permitted daily exposure
PK	Pharmacokinetics
PoD	Point of departure
PQRI	Product Quality Research Institute
QIVIVE	Quantitative *in vitro-in vivo* extrapolation
QSAR	Quantitative structure activity relationship
QWP	Quality working party
RAX	Read-across
ROC-AUC	Receiver operating characteristic area under the curve
RP	Reflection paper
SAR	Structure activity relationship
SWP	Safety working party
TK	Toxicokinetic
TTC	Threshold of toxicological concern

38

1. Executive summary

The ICH Q3A and Q3B guidelines provide a framework for qualifying Non-Mutagenic Impurities (NMI) in drug substances and products but offer limited guidance on new or elevated impurity levels. The reflection paper recognises the need for an adequate safety evaluation and suggests alternative strategies to *in vivo* animal studies for qualifying novel impurities. Impurities may be qualified when these are also present as significant metabolites in animals or humans. Impurities similar to the Active Pharmaceutical Ingredient (API) are generally covered by existing toxicological studies.

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 exposure level, route of administration, physicochemical (PC) properties, bioavailability, degradability, clinical conditions, and target population. The Threshold of Toxicological Concern (TTC) is an effective risk assessment tool for low-level exposures.

If there is a need for data, the primary source should be existing toxicological data that can be used to derive an Acceptable Level (AL). The AL method estimates a product-specific safe level of exposure to impurities. It is based on the Permitted Daily Exposure (PDE) methodology described in ICH Q3C, but also considers bioavailability, read-across (RAX) data and product-specific considerations. It involves selecting a point of departure (PoD) from toxicological studies and applying assessment factors.
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 should *in vivo* studies be considered. This reflection paper provides specific recommendations for this situation.

Impurities in investigational medicinal products should be evaluated according to ICH M3(R2), with special attention to impurities of higher concern, as well as considering short-term treatment as a de-risking element. When there is a need for additional safety data, the principles in this reflection paper can be applied.

In summary, when impurity-specific safety information for NMI is required, alternative strategies to gathering this information may be followed, including the use of existing toxicological data, RAX, TTC, computational and *in vitro* approaches. This information can be used in an integrated risk assessment. A weight-of-evidence (WoE) approach that includes an 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

ICH Q3A and Q3B are the core ICH quality guidelines that address qualification of NMI. They state that "qualification is 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. The applicant should provide a rationale for establishing impurity acceptance criteria that includes safety considerations." For DNA reactive (mutagenic) impurities, elemental impurities and residual solvents, specific guidance is provided in ICH M7(R2), Q3D and Q3C, respectively. For NMI outside the scope of these guidelines little guidance is available on how these impurities should be qualified. This is especially true when novel impurities are identified that were not present in the drug substance or drug product batches used in non-clinical safety and/or clinical studies during development, or when a higher level of these impurities needs to be qualified.

According to ICH Q3A/B guidelines, the level of any impurity present in a new drug substance or drug product that has been adequately tested in safety and/or clinical studies would be considered qualified. The limitation of this approach is that only the biological safety of a drug substance or drug product with a given impurity profile has been established (i.e. qualified), which is not the same as characterising the safety profile of an impurity. When toxicity is observed, it is usually not possible to discriminate between toxicity attributable to the active pharmaceutical ingredient (API) and toxicity attributable to the impurities present in the drug substance or drug product batch. This contrasts with the approaches developed for mutagenic impurities, elemental impurities and solvents, as well as the approach under development for extractables and leachables (ICH Q3E), where a compound-specific approach is recommended.

The safety testing of a drug substance only establishes that a drug substance batch with a certain impurity profile has a specific safety profile. This limits the possibility of extrapolating the safety of a drug substance, or product with a given impurity profile to a drug substance or product with the same API, and an increased level of one (or more) of its impurities, when no impurity-specific data are available. Also, when new impurities arise due to manufacturing changes or novel degradants are discovered, and these impurities cannot be controlled at a level below the qualification threshold, a lack of impurity-specific safety data complicates the qualification process.

The ICH Q3A/B guidelines recommend that additional safety testing should be considered when higher levels or new impurities need to be qualified. Consequently, for many impurities dedicated animal studies have been performed with the goal to qualify these impurities. Yet, concerns have been expressed from a scientific and a 3Rs perspective. Impurities in drug substances/products are dosed in these animal studies at much lower levels than the API. A No Observed Adverse Effect Level (NOAEL) in these studies is more likely related to the API instead of the impurities present (Graham *et al.*, 2021). This compromises the scientific rationale for the design of the qualification study and is in violation of Directive 2010/63/EU on the protection of animals used for scientific purposes, as from a 3Rs perspective, no animal studies should be performed if these studies are unlikely to provide relevant information. In fact, a survey amongst stakeholders reviewed *in vivo* studies on 467 impurities, which did not provide any examples where toxic impurities were identified. In these studies, in 98.7% of the cases the impurity was present, either spiked or unspiked, at a low level in the drug substance (Slikkerveer *et al*., 2024).

In case impurities have not already been qualified in previous safety studies (i.e. novel impurities) or when higher levels of impurities need to be qualified (that were previously qualified at a lower level), it is recommended to use alternative methodologies. Only in rare cases where a remaining concern cannot be resolved otherwise, should the conduct of an animal study be considered. In those cases, only a study with the neat impurity might provide relevant information on the safety profile of the impurity.

This reflection paper replaces a previous draft version (Reflection paper on the qualification of non-genotoxic impurities: EMA/CHMP/SWP/545588/2017). The focus of this reflection paper is to provide alternative strategies to qualify novel impurities or to qualify higher levels of impurities that were previously qualified at a lower level. It considers that the level of concern for impurities may vary depending on many factors that determine how much data is needed, ranging from none to compound-specific experimental data.

This reflection paper addresses the qualification of NMI, meaning the acceptability of certain levels of NMI from a safety perspective. It does not consider the acceptability of impurity levels from a quality perspective.

3. Scope

This paper reflects on the product-specific qualification of NMI in chemically synthesised pharmaceuticals.

Guidance on specific classes of impurities is provided in separate guidelines. These specific classes of impurities which are excluded from the scope of this reflection paper are:

- solvents (ICH Q3C),
- elemental impurities (ICH Q3D),
- extractables/leachables (ICH Q3E, under development),
- impurities in oligonucleotides (Guideline on the Development and Manufacture of Oligonucleotides (EMA/283093/2016)),
- 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)).

Impurities in Advanced Therapy Medicinal Products (ATMP), in herbal medicinal products, and in biological and biotechnologically derived pharmaceuticals are out of scope, although for products where both chemically synthesised and biotechnologically derived moieties are present (e.g. antibody drug conjugates) the same principles as described in this reflection paper may be considered.

The principles and methods discussed in this reflection paper should primarily be considered for the qualification of novel impurities arising from changed manufacturing processes, discovered after safety studies have been concluded, or when higher levels need to be qualified and existing data from safety studies are not sufficient for qualification. In accordance with ICH Q3A/B, the level of any impurity present in a new drug substance or drug product that has been adequately tested in safety and/or clinical studies is considered qualified.

Impurities present in products in clinical development are not in scope of ICH Q3A/B. See however, section 4.8. 4.8. In clinical trial approval procedures, the qualification of impurities has been a matter of debate and, in lieu of specific guidance, this reflection paper will discuss how the principles and methods described can be of help when considering the potential increased risk for clinical trial participants due to the presence of (novel) impurities.

4. Key considerations

4.1. General outline for risk assessment of NMIs

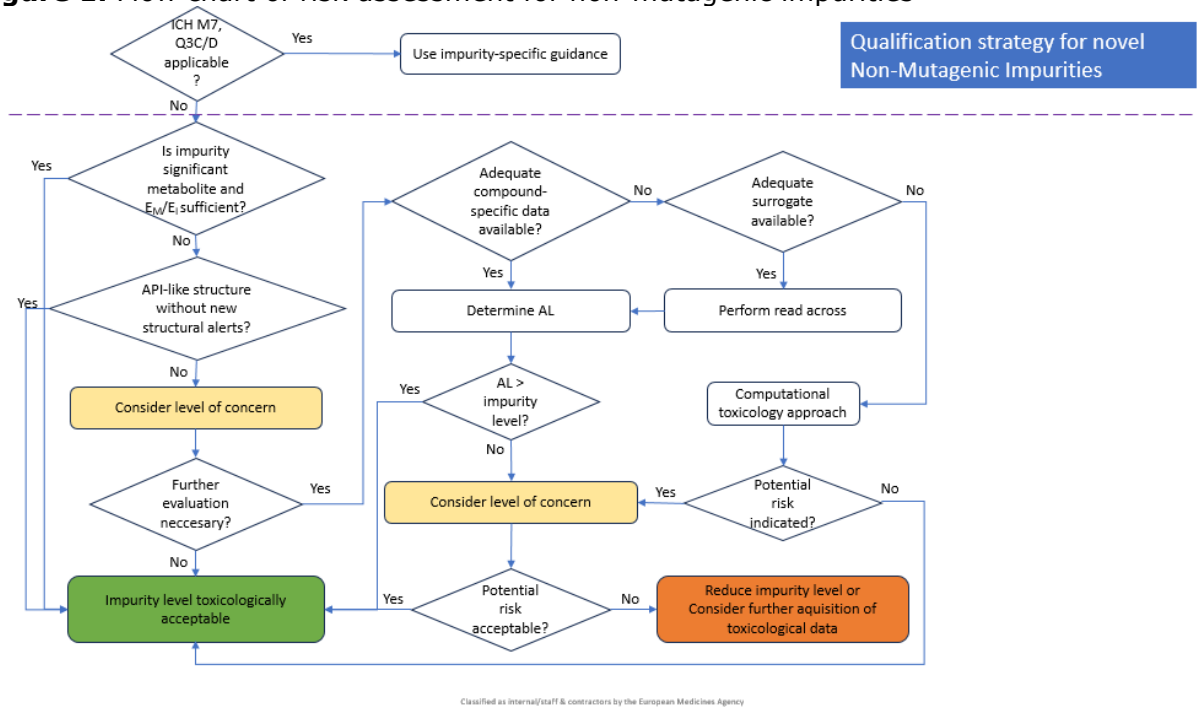
For NMIs, the general strategy for risk assessment is outlined in Figure 1. In the first step of toxicological qualification of a novel impurity, the applicability of this reflection paper should be checked. If no other guideline applies (e.g. ICH M7(R2), Q3C, Q3D), the principles outlined in this reflection paper should be considered for toxicological qualification of the impurity.

Impurities that are also metabolites in animals and/or humans may be qualified based on studies conducted with the API in relevant species (see section 4.2 Metabolites). In general, API-derived molecules without new structural alerts are considered of low toxicological concern and they are considered covered by the toxicological studies of the API (see section 4.3. API-like vs. non-API-like impurities). For all other molecules, the level of toxicological concern (see section Level of concern considerations) should be evaluated, e.g. is the new chemical structure/alert similar to molecules/structures with well-known toxicological profiles? (see section 4.5.1 Read-across). When there is no toxicological concern at the specified level of the impurity, no further evaluation is necessary. In case the level of concern (see section 4.4 Level of concern considerations) requires further evaluation of the toxicological profile of the impurity, the evaluation should start with searching for already available compound-specific data relevant for toxicological risk assessment and derivation of an Acceptable Level (AL) accordingly (see section 4.6 Acceptable Level calculation). If the level of the impurity is below the AL, the impurity is considered toxicologically qualified. If the level of impurity is greater than the AL, the level of concern (see section 4.4 Level of concern considerations) should be reconsidered taking into account the toxicological properties of the impurity, and the extent by which the AL is exceeded. If no compound specific data are available, an evaluation for adequate surrogate molecules with relevant compound specific data should be performed. In case an adequate surrogate molecule with relevant data can be identified, a RAX-approach (see section 4.5.1 Read-across) and use of the surrogate data to derive an AL is recommended. When it is not possible to identify an adequate surrogate molecule, it is recommended to start with computational toxicology approaches (see section 4.5

New approach methodologies) to identify and evaluate the toxicological potential of the molecule. Depending on the potential for toxicity identified and the uncertainties of the computational models available, a level of concern evaluation on potential risks, also considering maximum exposure levels, is recommended. If the potential risk is considered acceptable, no further data may be necessary. If the level of concern points to a potentially unacceptable risk at the maximum daily exposure, the impurity level should be lowered, or further toxicological data should be acquired. In this instance *in vitro* approaches may be considered (see section 4.5

New approach methodologies). Only in exceptional circumstances when acquisition of relevant data is only possible in *in vivo* studies, should conduct of *in vivo* studies be considered (see section 4.7. In vivo qualification studies).

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



4.2. Metabolites

According to ICH Q3A and Q3B, impurities that are also significant metabolites present in animal and/or human studies are considered qualified. If sufficient levels of the metabolite have been demonstrated, these impurities can usually be qualified using existing non-clinical studies with the API. The term "significant metabolite" should not be understood in the context of ICH M3(R2), which defines the threshold of 10% for when additional safety data on a metabolite is required. Non-clinical and clinical data for metabolites present below 10% are also considered relevant for qualifying an impurity. Rather, it is related to the extent of exposure to the metabolite (E_M) relative to the exposure level of the impurity (E_I). As shown in the flow chart (Figure 1), the ratio (E_M/E_I) of the exposure to the metabolite and the anticipated exposure to the impurity should suffice. For a metabolite/impurity of low concern, it may be sufficient when unity is reached. However, when potential toxicity of the impurity (metabolite) is of concern, a larger E_M/E_I needs to be considered.

While animal metabolism data is preferred for qualifying an impurity, metabolism data in humans can also be used. 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 drug product specification relative to the maximum daily dose (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 and to compare this with the maximum observed concentration (MOC) of the metabolite. The average plasma concentration C_{max} in the relevant dose group of animals or patients/volunteers would be considered to represent the MOC.

When calculating the MTC, worst case assumptions are considered, such as complete bioavailability with no plasma-protein binding, no distribution onto 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, 2022). The MOC of the metabolite is taken from an animal toxicity study or a relevant clinical study, preferably at a NOAEL or at clinically relevant exposure levels. Assuming these worst-case considerations for the systemic exposure of a small molecule impurity, using the MOC/MTC ratio for assessing the adequacy of the exposure margin between the levels of impurity and metabolite is considered a pragmatic approach.

4.3. API-like vs. non-API-like impurities

Impurities that are structurally similar to the API and retain the majority of functional groups, can be considered API-like, e.g. clobetasol propionate, clobetason-17-propionate and betamethasone. This may include degradation products of oxidation or hydrolysis reactions, where minor modifications to the structure are introduced, and the difference does not affect the overall structure or size of the molecule. 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 dimerization bridge does not introduce a new toxicophore (e.g. empagliflozin sugar dimer) and the impurity will degrade back into the parent compound once it has entered the systemic circulation. The term toxicophore is defined in section 4.5.1.1. 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 below, similarity may be evaluated by using computational predictive tools that identify toxicophores and predict physicochemical (PC) and pharmacokinetic (PK) properties for the impurity, to compare the properties between the API and the impurity. When it is concluded that no new toxicophores are introduced in the impurity compared to the API, and that PC and PK parameters are not significantly affected and therefore not likely to increase toxicity, the impurity is considered API-like. Consequently, no further investigations are required, as the toxicological properties of the impurity are covered by existing non-clinical studies with the API.

When impurities do show significant structural differences with the API, or where significant differences in PC 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.

Furthermore, there are exceptions where a non-significant change of the chemical structure would considerably change the toxicological potency of a molecule. As in medicinal chemistry, such activity cliffs can be relevant for toxicity, especially where a key event is based on binding of the compound 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).

4.4. Level of concern considerations

Various aspects related to the impurity at hand may affect the level of concern, which is reflected in Figure 2 as an outline for performing an assessment of the level of concern of a given impurity. Figure 2 is not suitable as a tool to characterise risk in a quantitative manner. It aims to support an integrated view on the level of concern in qualitative way. Foremost the level of exposure to the

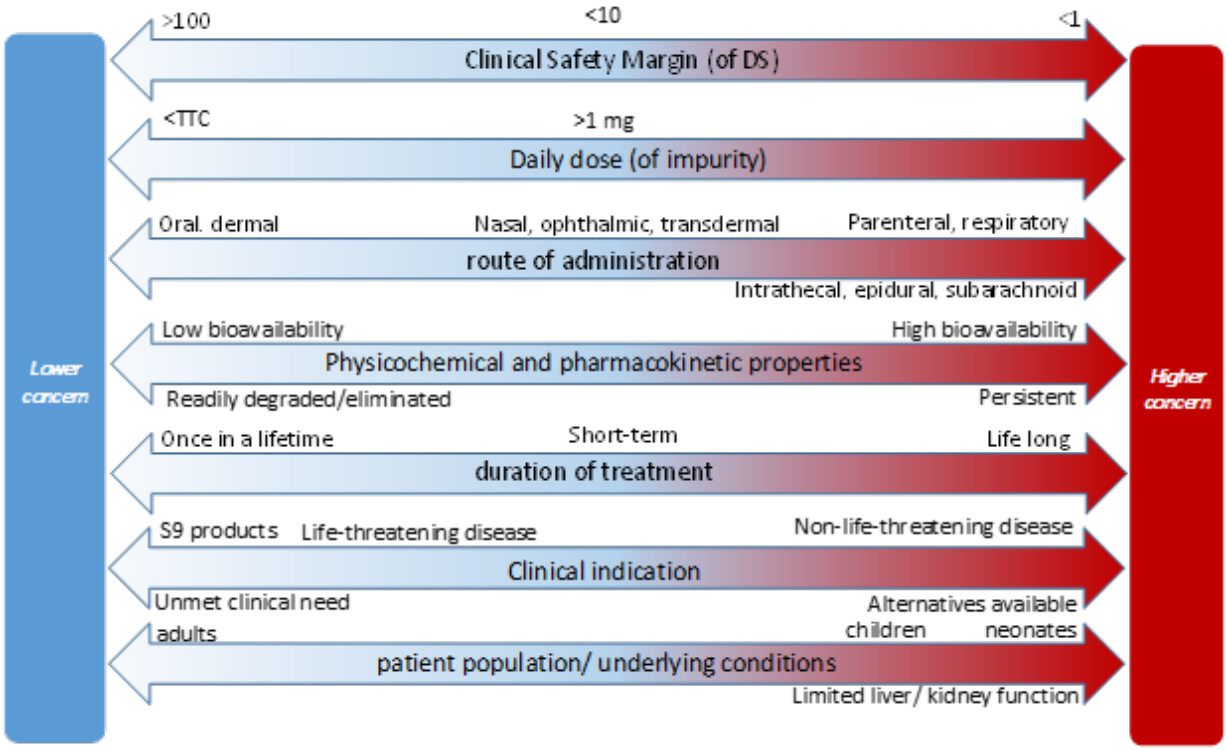
impurity needs to be considered, but also the route of administration can be of relevance. In addition, clinical conditions need to be considered such as the duration of treatment, indication and target population. Weighing these factors can support a conclusion on the level of concern. The overall level of concern should be considered to decide on further actions. Weighing the risk factors during evaluation of an impurity may obviate the need for further evaluation. If literature data is gathered or an *in-silico* assessment is performed, and a potential safety concern is noted, this information may be viewed in the context of the level of concern considerations and may trigger a decision to generate more (experimental) data to conclude on the limit for the impurity. Each risk factor needs to be considered in the context of all other aspects.

4.4.1. Exposure level considerations

4.4.1.1. Clinical safety margin of the drug substance

When impurities are present in drug substance batches that have been evaluated in regular safety studies, the NOAEL is usually determined by the toxicological profile of the API and not of the impurity. Nevertheless, a large safety margin (i.e. the ratio between the exposure in the animal study at the NOAEL and the maximal clinical exposure) would imply that both the API and the impurities present pose little toxicity at the tested levels. Consequently, this information can be considered for those cases where the impurity was already evaluated in (non-)clinical safety studies.

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



4.4.1.2. Absolute daily dose of impurity

The absolute daily dose or exposure to an impurity is an important consideration when weighing the perceived risk. In Figure 2, it is shown that a daily dose of an impurity at or below the Threshold of Toxicological Concern (TTC) is considered to be of low concern, which will be further explained below. In the regulatory context for pharmaceutical impurity assessments, a limit of 1 mg/day has been in use as qualification threshold for drug substances with a MDD of 667 mg/day or above (ICH Q3A). Assuming a body weight of 50 kg, this is equivalent to 0.02 mg/kg bw/day. In recent papers it has

been argued that APIs with a NOAEL below 0.02 mg/kg bw/day are extremely rare (Graham *et al*., 2021; Harvey *et al*., 2017; Kenyon *et al*., 2024; Slikkerveer *et al*., 2024). In these papers, the stance is taken that this would support 1 mg/day as a safe limit for impurity exposure. For most chemicals of low concern, exposure levels below 1 mg/day are considered to be safe. However, as there are certain chemical classes and chemotypes known to be associated with toxicities at lower exposure levels (Harvey *et al*., 2017), a safety concern cannot be excluded *a priori*. In addition, it also needs to be considered that inter- and intraspecies differences in sensitivity may occur. In addition, animal toxicity studies may be of shorter duration, whereas administration to patients can be long-term. For these reasons, a 1 mg/day exposure level is not used as a definitive cut-off value; instead, it has been included as the middle point in the daily dose bar. So, when exposure levels of impurities are above TTC and below 1 mg/day the level of concern needs to be evaluated taking into consideration all other aspects as shown in Figure 2. Furthermore, bioavailability differences between different routes of administration needs to be considered as discussed below. If the daily exposure to the impurity were to increase further, then it follows the concern would increase in parallel.

Threshold of Toxicological Concern

The TTC is a risk assessment tool for evaluating low-level exposure to chemicals with limited toxicological data. As described in ICH M7(R2), for mutagens the TTC is well-established within the pharmaceutical field. For non-mutagenic endpoints, the same principle can be applied when insufficient toxicological data is available for a single compound and no adequate RAX is possible (see section 4.5.1 Read-across). As a TTC is defined for a specific endpoint and route of exposure, the applicability of using a specific TTC needs to be verified. If the exposure level is below the relevant TTC, there is no need for further action. As TTC levels represent threshold levels for which there is no safety concern for most, but not all, chemicals, the level of concern still needs to be considered in the context of all other aspects as shown in Figure 2, even when the exposure to the NMI is below the TTC level.
Within the food area, for organophosphates or carbamates, the relevant TTC value is 0.3 µg/kg bw/day. Other substances are grouped according to the Cramer classification. The TTC values for Cramer Classes I, II and III are 30 µg/kg bw/day, 9 µg/kg bw/day and 1.5 µg/kg bw/day, respectively (EFSA, 2019). These values are based on the original work of Munro *et al*. (1996). For orally administered drugs, these values could be used as TTC when evaluating impurities, provided that the compound is not exempted from the application of the TTC because the compound is out of the applicability domain or has special properties such as steroids or bio-accumulative properties. Using various sources and refined methodologies, alternative values have been produced, which are generally in good concordance with the values cited above (e.g. Tluczkiewicz *et al*., 2011). Further refinement of the classification of compounds may become available in the future, which could justify the use of modified TTC values, e.g. when the results of the 'The Expanded Decision Tree (EDT) Project' by the US Food and Drug Administration (FDA) are available (Stice *et al*., 2021).
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*., 2017; Nelms and Patlewicz, 2020). Cramer classification appeared to be a less suitable approach for this route. For the compounds 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 qualification threshold of 5 µg/day for leachables in orally, inhaled, and nasal drug products as derived by the PQRI consortium (Ball *et al*., 2007). The latter mentions that irritating compounds, 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 drugs 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 sensitization thresholds (DST) have been calculated for non-reactive, reactive and HPC chemicals, which are 710, 73, and 1.0 µg/cm^2, respectively (Parris *et al.*, 2023; Chilton *et al*., 2022). These values could be used for dermal products, to address the concern for dermal sensitisation. For other non-mutagenic endpoints, the systemic exposure needs to be considered, taking into account the surface to which the product is applied and the degree of dermal absorption (see below 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, 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 by combining the oral NOEL values with available PK data, or where such data were not available, by applying 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 that allow the derivation of a parenteral PDE from an oral PDE, 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 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 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. Another effort to establish a parenteral qualification threshold for extractables and leachables is on its way through the ICH Q3E expert working group (EWG), using an extended Permitted Daily Exposure method. Pending the results of this EWG and considerations regarding the extrapolatability of these results for extractables and leachables to pharmaceutical impurities in general, 5 µg/day appears to be a sufficiently protective value for a TTC to apply to pharmaceutical impurities administered parenterally. The TTC and DST values that can be used for NMI are summarised in route of administration.

4.4.1.3. Route of administration

Besides route-dependent differences in bioavailability, which are discussed below, the route-dependent differences in toxicity need to be considered. These route-specific sensitivities are also reflected in the different TTC values as discussed above. For instance, orally inhaled and nasal drug products are delivered to the respiratory tract where tissues are receptive to sensitisation and irritation. Consequently, these endpoints are often the most critical endpoints 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. 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 sensitivity of central nervous system (CNS) tissues needs to be considered for these special routes.

Table 1. 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 Dermal sensitisation threshold (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). HPC = High Potency Category[4] for other non-mutagenic endpoints.

4.4.1.4. Physicochemical and pharmacokinetic properties / bioavailability

By definition, compounds administered via a parenteral route have 100% bioavailability, and consequently, they pose the highest concern, as opposed to compounds administered via routes where limited absorption may reduce the systemic exposure. 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 may be retrieved from the literature. In the absence of such data, PC properties can be considered to estimate bioavailability. These properties are also used in *in silico* tools to estimate bioavailability. As these tools have their limitations, predictability can be improved by supporting experimental New Approach Methodologies (NAM) data such as transport across Caco-2 cells and metabolism in hepatic models. In case of (dia)stereoisomers, potential effects on PK properties need to be considered (Section API-like vs. non-API-like impurities). In the absence of factual data to the contrary, bioavailability of compounds administered via the respiratory route is considered to be (close to) 100%.

Compounds that are poorly degraded or eliminated otherwise increase the level of concern as such compounds can accumulate and, even with low daily exposures, may reach tissue concentrations where adverse effects could occur.

4.4.2. 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 NMI, especially for sensitive populations such as children, patients with renal or hepatic impairments, and pregnant women, as well as for severe pathologies where the benefit-risk balance differs. This should also take into consideration the duration of treatment.

4.4.2.1. Duration of treatment

The treatment duration is a key factor to consider when determining the level of concern for impurities in pharmaceutical products. For short-term treatments, due to the brief exposure of the patient to the impurity, the level of concern for impurities is usually lower: a single dose, treatments lasting less than a month or intermittent dosing result in reduced cumulative exposure, thus mitigating long-term health risks.

Conversely, chronic treatments, particularly those that last throughout a patient's lifetime, necessitate a more thorough evaluation of impurity levels due to the increased cumulative exposure, which increases the potential risk from impurities. Therefore, where treatment duration increases, the level of concern does, too.

4.4.2.2. Clinical indication

The clinical indication should be considered as a critical factor in the determination of acceptable levels of NMI in drug substances, as stipulated in the ICH Q3A and Q3B guidelines. These guidelines provide the primary criteria for qualifying NMI, but they also permit modifications to the qualification thresholds, either upward or downward, depending on the medicinal product involved. Such modifications are founded on a scientific rationale encompassing clinical indications and the related level of concern.

In the context of severe or life-threatening diseases or products with a high clinical need, the presence of impurities may be justified due to the imperative need for therapeutic options. Nevertheless, as an integral part of the risk-benefit assessment process, any acceptance of increased impurity levels must be scientifically substantiated and confined within pre-established safety parameters. The ICH S9 guideline specifically addresses the management of impurities in anti-cancer medications, noting that the imposition of impurity controls identical to those applied to less severe conditions is inappropriate owing to a different risk-benefit consideration. Furthermore, alterations in the clinical applications of marketed products, such as the introduction of new indications for less severe conditions, may necessitate the re-assessment of existing impurity specifications to ensure continued compliance with safety standards.

Therefore, the clinical indication, encompassing the severity of the condition and the risk-benefit analysis of the treatment, plays a pivotal role in determining whether the standard impurity limits remain suitable or whether modifications are warranted based on the specific clinical application of the drug. To safeguard patient safety and treatment efficacy, it is imperative that impurity assessments are customised for each pharmaceutical product and its intended therapeutic use.

4.4.2.3. Target population

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

- Paediatric populations: children, especially infants and young children, are more susceptible to the toxic effects of impurities due to differences in absorption, distribution, metabolism, and excretion compared to adults.
- Patients with renal or hepatic disease: conditions such as renal or hepatic failure can lead to increased concentrations of impurities due to impaired elimination. Products for these patients may require specific evaluations of impurity thresholds to ensure patient safety.
- Pregnant individuals: exposure to impurities during pregnancy is concerning due to potential developmental toxicity to the fetus. Although specific data on NMI is limited, pregnant individuals generally represent a population that necessitates careful risk assessment.

4.5. 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.

4.5.1. Read-across

If no sufficiently robust data can be identified on the impurity itself, it is possible to perform RAX to one or more surrogate compounds for which robust data is available or by using a grouping methodology, to identify qualitative or quantitative data used for qualifying the impurity at the specified level.

4.5.1.1. Surrogate approach

When performing RAX to a surrogate compound, firstly, the impurity should be characterised in terms of chemical-structural properties as well as PC and PK properties.

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 compound. This can include both pharmacologically active and non-active moieties of the compound. PC properties (such as polarity, solubility, lipophilicity, ionizability, and molecular weight), as well as PK properties (such as bioavailability, distribution, metabolism, and excretion) should be presented, e.g. from databases or based on predictions using computational tools. When *in silico* tools are used, it should be justified these are fit-for-purpose (see section New approach methodologies). 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 impurity. 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 close to the toxicophore, which could potentially affect the biological activity, should be identified. The global chemical similarity could also be assessed, and e.g. expressed by the Tanimoto score. Comparability based on PC or PK properties should be discussed. The choice of adequate surrogate(s) should be justified based on the similarity and uncertainties with the RAX method and the adequacy of the outcome of the assessment should be provided together with the overall outcome of the RAX approach.

As detailed in the Computational toxicology section, 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 PC 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. Nonetheless, computational tools could be used for identifying toxicophores considered relevant for major targets (liver, kidney, cardiovascular system (CVS), gastro-intestinal tract (GIT), CNS and respiratory system (RS)). Applicants are encouraged to gather more data for qualification of new endpoints in (Q)SAR predictions, which could further increase the validity of prediction tools for specific endpoints.

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

4.5.1.2. Grouping approach

Alternatively, a grouping approach can be used, where several similar compounds are grouped, containing the same toxicophores and functional groups, which allows the detection of trends across endpoints. Again, as defined for the RAX approach, adequate similarity must be demonstrated for chemical-structural, PC and PK properties to group the compounds. Based on the assessment of similarity of the grouped compounds, 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 compounds is not likely to present any safety risk at the specified level of the impurity.

4.5.2. Computational toxicology

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. A wide range of methods including (Q)SAR, Machine Learning (ML) models and Artificial Intelligence (AI) are part of the NAMs. These *in silico* methods are usually based on broad databases and training sets of chemical compounds tested *in vitro* and *in/ex vivo*.

If the methods outlined above have not resulted in relevant data for qualifying the impurity, computational predictive *in silico* tools can be used to identify potential safety alerts i.e., toxicophores of the impurity, and to further characterise the safety concern (see section Hazard characterisation and quantitative risk estimation).

4.5.2.1. Choice of the tool

In addition to traditional (Q)SAR and RAX approaches (see section above), AI/ML methodologies and potentially Adverse Outcome Pathways (AOPs) are already available. It is possible to rationally combine evidence from several *in silico* tools or both *in silico* and *in vitro* studies to fill in knowledge gaps regarding toxicological events, for example by using the AOP approach, which offers a framework for organising data at the chemical and biological levels.

Any *in silico* tool used in the risk assessment should be justified by the applicant and complemented by expert knowledge opinion in the Expert review. This opinion should also summarise the applicability of the tool 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. Important guidance is given for example by the five OECD principles (OECD, 2004) and the OECD Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Model, 2007 and by OECD (Q)SAR Assessment Framework, 2023. In general, the best practices available 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)

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. The absence of the complementary method for the chosen endpoint(s) should be justified.

4.5.2.2. 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, CNS and RS), ideally in terms of functional and tissue organ changes. Models for the PC and PK properties can be used to further define risks associated with the impurity along with identified toxicophores. An alert for non-mutagenic carcinogenicity should also be considered and human relevance for mode of action (MoA) 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 models' 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; 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.

4.5.3. *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 compounds 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 microphysiological systems) with careful selection of endpoints may be considered. 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 RAX 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 API to test systems would complicate the interpretation of the read-outs, as the API itself may also have an effect in the *in vitro* model employed.

4.5.4. Hazard characterisation and quantitative risk estimation

Foremost, NAM tools provide qualitative data. To use NAM tools to estimate safety risks quantitatively, more data are needed, especially to translate quantitative *in vitro* data to the *in vivo* situation (QIVIVE). Therefore, the current application of NAM tools is often focused on hazard characterisation (Schmeisser *et al*., 2023). When NAM tools indicate the absence of relevant hazards for major targets (liver, kidney, CVS, GIT, CNS and RS), this information can be included in the weight of evidence approach for the safety assessment of the impurity. If a hazard is identified with a NAM tool, it needs to be demonstrated that the potency of the impurity to display the associated toxicity is not of concern. It may be sufficient to justify that the exposure to the impurity at the proposed limit is without safety concerns.

4.6. Acceptable Level calculation

Deriving an estimate of a safe level of exposure in patients by using toxicological data is generally achieved by choosing a Benchmark Dose Lower boundary (BMDL) or a NOAEL in a toxicological study and applying assessment factors to correct for variability, uncertainties and known differences between the animal model used and the patient for whom a predicted safe level is needed. Here we propose the Acceptable Level (AL) method, by which similar toxicological principles are used, e.g. as described in ICH Q3C and ICH Q3D. The AL is 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 safety. 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 with regard to the level of concern relevant for the product can be taken into account. 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). Subsequently, assessment factors (AF) 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).

Drug impurities cover a wide chemical space and consequently, 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 assessment factor is used to account for differences in bioavailability (AF6).

In this reflection paper the use of RAX is described as an alternative when insufficient compound-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 assessment factor is recommended (AF7).

The AL can be calculated with the formula:

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

The use of assessment factors is described in more detail in the Appendix.

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 which 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.
A 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. Furthermore, the AL method includes corrections for bioavailability and considers uncertainty related to a surrogate approach, whereas the PDE method does not.

4.7. In vivo qualification studies

4.7.1. Design of *in vivo* studies

Performing *in vivo* studies to qualify the toxicological properties of an impurity is generally discouraged in light of the 3Rs principles. Furthermore, *in vivo* studies that have been conducted for the qualification of new impurities as required by ICH Q3A, especially when the study investigated (impurity-spiked) batches of the API at dose levels at or below the NOAEL of the API have provided limited additional information. However, if all other alternative options mentioned above have not provided the necessary information to qualify an impurity at a proposed specification limit, a preferred *in vivo* study design is presented here, to harmonise the approach for deriving a PoD for setting an AL for impurities based on *in vivo* data.

Several industry-led publications have investigated the preferred study design among sponsors and have provided recommendations for an *in vivo* study design (Mitra *et al*., 2021; Slikkerveer *et al*., 2024). While some of the principles laid out in these publications are endorsed, others require further reflection in terms of optimal study design to ensure sufficient levels of exposure to the impurity. This is needed to establish adequate exposure margins to the proposed specification level of the impurity when taking the method for calculating an AL into account, as described in the section above.

As observed by both papers, the preferred test item among sponsors is stated as API batches spiked with the impurity to a specific level, in order to achieve a certain exposure margin. This design is flawed however, as the levels of impurities may not be sufficient to ensure an adequate ratio between the impurity level and the AL. Uncertainty regarding adequate exposure could trigger repetition of *in vivo* studies, which is not favoured considering 3Rs principles. It is therefore recommended to perform the *in vivo* study on neat samples of the impurity (i.e. isolated impurity without API) with a purity of > 95%, also to enable evaluating the effects of the neat impurity itself without the potentially added effects of the API. Moreover, the *in vivo* study should be GLP compliant and adhere to the principles of OECD test guideline 407. The typical duration of the studies was also investigated (Mitra *et al*., 2021, Slikkerveer *et al*., 2024), and 28 days of repeated dosing via the clinical route of administration was recommended. For medicinal products intended for short term administration, the duration of the study could be reduced to 14 days. No recovery period for the treated groups was recommended, but a vehicle control group should be included. The most used species for impurity testing is rat. Finally, TK should be included in the study; however, TK analysis can be integrated in the main study as part of the high dose group, so a separate TK group would not be necessary. Overall, these recommendations can be endorsed.

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, 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. In light of 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 rats/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. For the purpose of qualifying NMI, preliminary dose-range finding studies with the impurity are not recommended. Considering the usual assessment factors 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 F6 as additional assessment factor 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 2. 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 assessment factors.
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.

4.7.1.1. Special considerations for oncology products

Because impurities in cytotoxic oncology products may have similar toxic properties as the API, the approach described above may not be applicable. Instead, in line with ICH S9, qualification of impurities in these products may be based on the similarity in safety profile between the impurity and the API; there is no need to control the impurity at a level where no toxicity is anticipated.

4.7.2. Setting limits based on *in vivo* data

No dedicated guidance has been developed yet from EMA or ICH regarding the derivation of a BMDL for use as PoD as a basis for the AL. Until such dedicated guidance becomes available, the EFSA guidance from 2022 can be consulted on important principles. Based on the best fitted model and the most relevant BMDL, the AL can be derived. Further guidance on deriving the AL is given in section 4.6 'Acceptable Level calculation'.

4.8. Products under clinical development

An evaluation of impurities in investigational medicinal products under clinical development should be performed according to ICH M3(R2). The evaluation can be performed using relevant parts of the reflection paper as outlined above. Special consideration should be given to impurities of higher concern, e.g. based on prior knowledge of groups of impurities or specific chemical features, while impurities of lesser concern would not need to be investigated to the same extent. Impurities of high concern could lead to a request for lower batch levels or require more data to qualify the impurity. For evaluation of an impurity, the level of concern analysis should be applied to impurities in products under clinical development to direct the extent of further qualification, as detailed in the sections above. This evaluation also includes considerations of short-term treatment as a de-risking element, which can be a relevant aspect in the clinical trial setting. If the levels of impurities reported in the clinical trial batches are considered to be of sufficiently low concern, no further information needs to be submitted.

5. Conclusion

When impurity-specific safety information for NMI is required, alternative strategies to gather this information may be followed, including the use of TTC, (Q)SAR, RAX and *in vitro* approaches. This information can be used in an integrated risk assessment. A WoE approach that includes an all aspects that determine the level of concern, could be sufficient to decide that the NMI can be considered safe at the specified level

6. References

Arnot JA, Toose L, Armitage JM, Sangion A, Looky A, Brown TN, Li L, Becker RA. Developing an internal threshold of toxicological concern (iTTC). J Expo Sci Environ Epidemiol. 2022 Nov;32(6):877-884. doi: 10.1038/s41370-022-00494-x.

Ball D, Blanchard J, Jacobson-Kram D, McClellan RO, McGovern T, Norwood DL, 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 Jun;97(2):226-36. doi: 10.1093/toxsci/kfm058.

Blum J, Masjosthusmann S, Bartmann K, Bendt F, Dolde X, Dönmez A, Förster N, Holzer AK, Hübenthal U, Keßel HE, Kilic S, Klose J, Pahl M, Stürzl LC, Mangas I, Terron A, Crofton KM, 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 Jan;311(Pt 2):137035. doi: 10.1016/j.chemosphere.2022.137035.

Chilton ML, Api AM, Foster RS, Gerberick GF, Lavelle M, Macmillan DS, Na M, O'Brien D, O'Leary-Steele C, Patel M, Ponting DJ, Roberts DW, Safford RJ, Tennant RE. Updating the Dermal Sensitisation Thresholds using an expanded dataset and an in silico expert system. Regul Toxicol Pharmacol. 2022 Aug;133:105200. doi: 10.1016/j.yrtph.2022.105200.

Dik S, Ezendam J, Cunningham AR, Carrasquer CA, van Loveren H, Rorije E. Evaluation of in silico models for the identification of respiratory sensitizers. Toxicol Sci. 2014 Dec;142(2):385-94. doi: 10.1093/toxsci/kfu188. Epub 2014 Sep 19. PMID: 25239631.

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

EFSA Scientific Committee. Guidance on the use of the benchmark dose approach in risk assessment. EFSA journal 2022. https://doi.org/10.2903/j.efsa.2022.7584

Escher SE, Tluczkiewicz I, Batke M, Bitsch A, Melber C, Kroese ED, Buist HE, Mangelsdorf I. Evaluation of inhalation TTC values with the database RepDose. Regul Toxicol Pharmacol. 2010 Nov;58(2):259-74. doi: 10.1016/j.yrtph.2010.06.009.

Escher SE, 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 LS, Taboureau OT, Tolosa L, Vrijenhoek NG, van Vugt-Lussenburg BMA, 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 Mar;79:105269. doi: 10.1016/j.tiv.2021.105269.

Graham JC, Powley MW, Udovic E, Glowienke S, Nicolette J, Parris P, Kenyon M, White A, Maisey A, Harvey J, Martin EA, Dowdy E, Masuda-Herrera M, Trejo-Martin A, Bercu J. Calculating qualified non-mutagenic impurity levels: Harmonization of approaches. Regulatory Toxicology and Pharmacology 126 (2021) 105023.

Golden E, Macmillan DS, 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. doi: 10.14573/altex.1911261. Epub 2020 May 7. PMID: 32388570; PMCID: PMC11316520.

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 Mar;84:116-123.

Kenyon MO, Martin Matthew, Martin EA, Brandstetter S, Wegesser T, Greene N, Harvey J. Deriving acceptable limits for process-related organic impurities in medicinal 1 products – Durational adjustments. Regul. Toxicol. Pharmacol. 2024 Jun:150:105644. doi: 10.1016/j.yrtph.2024.105644.

Masuda-Herrera MJ, Bercu JP, Broschard TH, Burild A, Hasselgren C, Parris P, Ford LC, Graham J, Stanard B, Comerford M, Lettiere D, Erler S, Callis CM, Morinello E, Muster W, Martin EA, Griffin TR, 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 Sep-Oct;76(5):369-383. doi: 10.5731/pdajpst.2021.012693.

Mitra MS, Datta K, Hutchinson R, Nicolette JJ, Pettersen JC, Wegesser TC, Bercu JP. Harmonized 3Rs-based non-mutagenic impurity qualification study designs developed using the results of an IQ consortium survey. Regul Toxicol Pharmacol. 2021 Jun;122:104895.

Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G., 1996. Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of toxicological concern. Food Chem. Toxicol. 34, 829–867.

Nelms MD, Patlewicz G. Derivation of New Threshold of Toxicological Concern Values for Exposure via Inhalation for Environmentally-Relevant Chemicals. Front Toxicol. 2020 Oct 16;2:580347. doi: 10.3389/ftox.2020.580347.

Nishijo T, Api AM, Gerberick GF, Miyazawa M, Roberts DW, Safford RJ, 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 Nov;117:104732. doi: 10.1016/j.yrtph.2020.104732.

OECD (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions, https://one.oecd.org/document/ENV/CBC/MONO(2023)32/en/pdfPaixão P, Gouveia LF, Morais JA. 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 Jun 15;429(1-2):84-98. doi: 10.1016/j.ijpharm.2012.03.019.

Parris P, Whelan G, Burild A, Whritenour J, Bruen U, Bercu J, Callis C, Chilton ML, Graham J, Johann E, Johnson C, Griffin T, Kohan M, Martin EA, Masuda-Herrera M, Stanard B, Cruz MT, Nagao L. Sensitization Assessment of Extractables and Leachables in Pharmaceuticals: ELSIE Database Analysis. PDA J Pharm Sci Technol. 2023 Sep 15:pdajpst.2022.012811. doi: 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 Jun;89(6):941-8. doi: 10.1007/s00204-014-1287-6.

Roberts DW, Api AM, Safford RJ, Lalko JF. Principles for identification of High Potency Category Chemicals for which the Dermal Sensitisation Threshold (DST) approach should not be applied. Regul Toxicol Pharmacol. 2015 Aug;72(3):683-93. doi: 10.1016/j.yrtph.2015.03.001.

Rovida C, Escher SE, Herzler M, Bennekou SH, Kamp H, Kroese DE, Maslankiewicz L, Moné MJ, 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. doi: 10.14573/altex.2010062

Schirris TJ, Renkema GH, Ritschel T, Voermans NC, Bilos A, van Engelen BG, Brandt U, Koopman WJ, Beyrath JD, Rodenburg RJ, Willems PH, Smeitink JA, Russel FG. Statin-Induced Myopathy Is Associated with Mitochondrial Complex III Inhibition. Cell Metab. 2015 Sep 1;22(3):399-407. doi: 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 GEN, Kleinstreuer N, Leist M, Luijten M, Marx-Stoelting P, Poetz O, van Ravenzwaay B, Roggeband R, Rogiers V, Roth A, Sanders P, Thomas RS, Marie Vinggaard A, 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 Aug;178:108082. doi: 10.1016/j.envint.2023.108082.

Schneckener S, Grimbs S, Hey J, Menz S, Osmers M, Schaper S, Hillisch A, Göller AH. 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 Nov 25;59(11):4893-4905. doi: 10.1021/acs.jcim.9b00460

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 Jun:150:105647. doi: 10.1016/j.yrtph.2024.105647.

Stice S, T. B. Adams, R. Kolanos, A. Mattia. US FDA's Updates to the Cramer et al. (1978) Decision Tree: The Expanded Decision Tree. SOT 2021 Virtual session: Thresholds of Toxicological Concern: Reassessing the Basis and Expanding the Horizon. 22 March 2021.

Stumpfe D, Hu H, Bajorath J. Evolving Concept of Activity Cliffs. ACS Omega. 2019 Aug 26;4(11):14360-14368. doi: 10.1021/acsomega.9b02221.

Ta GH, Weng CF, Leong MK. In silico Prediction of Skin Sensitization: Quo vadis? Front Pharmacol. 2021 May 4;12:655771. doi: 10.3389/fphar.2021.655771. PMID: 34017255; PMCID: PMC8129647.

Teubner W, Mehling A, Schuster PX, 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 Dec;67(3):468-85. doi: 10.1016/j.yrtph.2013.09.007. Epub 2013 Sep 30. PMID: 24090701.

The five OECD principles: https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf

The QSAR assessment framework: https://www.oecd.org/chemicalsafety/risk-assessment/qsar-assessment-framework.pdfTluczkiewicz I, Buist HE, Martin MT, Mangelsdorf I, Escher SE. Improvement of the Cramer classification for oral exposure using the database TTC RepDose – A strategy description. Regul Toxicol Pharmacol. 2011, 61:340-350.

Tluczkiewicz I, Kühne R, Ebert RU, Batke M, Schüürmann G, Mangelsdorf I, Escher SE. Inhalation TTC values: A new integrative grouping approach considering structural, toxicological and mechanistic features. Regul Toxicol Pharmacol. 2016 Jul;78:8-23. doi: 10.1016/j.yrtph.2016.03.022.

Tobias A, Ballard BD, Mohiuddin SS. Physiology, Water Balance. [Updated 2022 Oct 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK541059/

Wang CC, Lin YC, Wang SS, Shih C, Lin YH, Tung CW. SkinSensDB: a curated database for skin sensitization assays. J Cheminform. 2017 Jan 31;9:5. doi: 10.1186/s13321-017-0194-2. PMID: 28194231; PMCID: PMC5285290.

Watford S, Ly Pham L, Wignall J, Shin R, Martin MT, Friedman KP. ToxRefDB version 2.0: Improved utility for predictive and retrospective toxicology analyses. Reprod Toxicol. 2019 Oct;89:145-158. doi: 10.1016/j.reprotox.2019.07.012. Epub 2019 Jul 21. PMID: 31340180; PMCID: PMC6944327.

Wei Z, Xu T, Strickland J, Zhang L, Fang Y, Tao D, Simeonov A, Huang R, Kleinstreuer NC, 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 Feb 28;6:1321857. doi: 10.3389/ftox.2024.1321857. PMID: 38482198; PMCID: PMC10933113.

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. (2020). Qualification of impurities based on metabolite data. Regul Toxicol Pharmacol. 110, 104524.

Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, Patlewicz G, Shah I, Wambaugh JF, Judson RS, Richard AM. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform. 2017 Nov 28;9(1):61. doi: 10.1186/s13321-017-0247-6. PMID: 29185060; PMCID: PMC5705535.

7. Appendix

Use of assessment factors

The use of modifying factors for the derivation of a PDE is described in ICH Q3C/D documents. Here, we reflect on the use of assessment factors, 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 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 non-mutagenic impurities. It is difficult to justify that there are 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 modifications can be considered on a case-by-case basis, for instance to take into account that the product is administered intermittently, or only short-term (*i.e.* up to 1 month). In the case of drugs 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 drug product.

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. Choosing a value for AF4 depends not only on the severity of the effects observed, but also on the ratio between the dose level at which the severe effect is observed and the dose level chosen as PoD. For instance, when no BMDL or NOAEL is available and a LOAEL is used at which the severe effect is observed, a higher AF4 value (up to 10) should be considered. When a NOAEL is available and used as PoD, while severe toxicity is observed at the

LOAEL, the AF4 value could be more moderate. Yet the degree of moderation would depend on the spacing between dose levels, the severity of the effects at the LOAEL, and the steepness of the dose-severity curve.
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 with an example as below.

Table 3. 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 critical endpoint, the PoD would be 1 mg/kg/day and AF4 would be 1. When choosing the CNS toxicity as critical endpoint, the PoD would be 3 mg/kg/day and AF4 would be 10.

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 BMDL or NOAEL was not established

ICH Q3C indicates that a factor of up to 10 could be used depending on the severity of the toxicity. ICH Q3D differentiates between NOAEL/LOAEL and NOEL/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. Furthermore, the steepness of the dose-response curve is relevant in choosing the values, not only for AF4, but also for AF5. This may complicate the choice of the values for AF4 and AF5.

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 observed. 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. Using the example provided in Table 3, if 10 mg/kg/day were the lowest dose tested where severe CNS effects are observed, both AF4 and AF5 would be 10.

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 suggest 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:

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 RAX approach, these data may be leveraged to inform the bioavailability estimate, if sufficiently justified.

When the data concern 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 physicochemical 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 drug 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 RAX 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.