Meeting Report

Workshop Report on the Evaluation of the Updated and Expanded Carcinogen Database to Support Derivation of Threshold of Toxicological Concern Values for DNA-Reactive Carcinogens

doi:10.14573/altex.2210111

1 Introduction

Threshold of toxicological concern (TTC) values are frequently used for compounds occurring at low concentrations in feed and food or as impurities in drugs. This workshop report addresses the emerging alternatives for deriving TTC values for DNA-reactive carcinogens and evaluating the acceptability of the Cramer Class TTC values to be adequately protective for non-DNA-reactive carcinogens.

TTC values define a daily lifetime exposure limit below which an adverse effect on human health is not to be expected. They are used to assess low-level exposure to compounds lacking sufficient toxicity data *per se* and thus contribute to reducing animal testing, which is in line with the 3R principle (Russell, 1999). Even today there are many substances whose toxicity is widely unknown, like metabolites, degradation products, impurities, or process intermediates. In many of these cases, *in vivo* studies are not practical and/or technically not feasible due to the small quantities of the substances or because rapid decisions must be made, e.g., in the case of impurities in drugs that result from unintentional formation such as during their production or storage. TTC values are used, for example, to assess genotoxic impurities in drugs according to the ICH M7 guideline[1] or for compounds occurring at low levels in feed and food (EFSA, 2012).

While the TTC concept has been thought of as making a distinction between carcinogenic and non-carcinogenic compounds, the tiered decision tree makes a distinction based on structural alerts (and possibly data) relating to DNA reactivity and genotoxicity as surrogates for carcinogenic potential via a mode of action considered to have no threshold.

Chemical carcinogens have been broadly categorized as causing cancer by either a genotoxic/mutagenic or a non-genotoxic mode of action:

DNA-reactive (genotoxic) carcinogens are substances that have the capacity to cause direct DNA damage (such as DNA adducts or DNA lesions) at low exposure levels and in general are considered to not exhibit a biological threshold. Preston and Williams (2005) list 10 key events characteristic of direct DNA-reactive carcinogens. DNA-reactive carcinogens often form reactive electrophiles upon metabolic activation. These electrophiles bind covalently to nucleophilic cellular macromolecules, including genomic DNA, forming DNA adducts and/or produce other DNA lesions (Kobets and Williams 2019). Such DNA damage can result in gene mutations, which, if not repaired sufficiently, can lead to cancer. This type of genotoxic compound is usually detectable in a bacterial reverse mutation assay (i.e., Ames test, ICH guideline[1]). In addition to direct gene mutation, genotoxicity (or mutagenicity) includes clastogenicity, such as sister chromatid exchanges, chromosomal aberrations, or DNA strand breaks, and aneugenicity. In general, it has been accepted than aneugenicity is not of concern below the TTC for non-cancer effects. Most also apply a similar consideration to clastogenicity (e.g., EFSA and WHO, 2016; EFSA et al., 2019; WHO, 2020), while some consider that an assessment of clastogenicity is also necessary to decide on which TTC threshold is appropriate (EFSA and WHO, 2016). In general, clastogens also tend to be DNA-reactive, and hence the primary focus has been to exclude the possibility of DNA-reactive mutagenicity.

Non-DNA-reactive (non-genotoxic) carcinogenicity comprises several mechanisms that do not involve direct reaction of the chemical with cellular DNA or direct DNA damage. These mechanisms include immunosuppression, hormonal perturbation (direct mitogenesis), growth promotion, or chronic cytotoxicity followed by regenerative hyperplasia, oxidative stress, and other effects that result in increased cell proliferation and thereby in tumor increase. Such compounds are considered to exhibit a biological threshold in their dose-response.

The TTC values for non-genotoxic compounds are based on a large database of noncancer toxicity data on organic chemicals, which were classified based on a presumption of toxicity rated as low (Class 1), moderate (Class 2), or toxic (Class 3) using the Cramer decision tree (Cramer et al., 1976; Munro et al., 1996). The 5th percentile no observed adverse effect level (NOAEL) per class was used to derive a human threshold, taking into account an assessment factor of 100 to cover inter- and intra-species differences and other factors, where necessary, to extrapolate from a lowest observed adverse effect level or from a subchronic study.

[1] https://www.ema.europa.eu/documents/scientific-guideline/ich-guideline-m7r1-assessment-control-dna-reactive-mutagenic-impurities-pharmaceuticals-limit_en.pdf

The approach to derive acceptable exposure limits for potential DNA-reactive carcinogens is completely different. The TTC value for such compounds is based on an analysis of the potency for 730 carcinogens, largely those compiled in the Cancer Potency DataBase (CPDB) (Kroes et al., 2004). Based on the regulatory default assumption that there is no threshold for (DNA-reactive) carcinogenicity, the exposure limit is based on linear extrapolation of the dose at which 50% of tested animals developed tumors (TD50 value) down to an acceptable theoretical human risk of one in a million ($1/10^6$). In this conservative approach, the TD50 value per chemical is derived for the most sensitive species and tumor site (Cheeseman et al., 1999). Based on this analysis and a number of other considerations, a threshold of 0.15 µg/p/d, corresponding to 0.0025 µg/kg bw/d was proposed (Kroes et al., 2004). Since then, a number of questions have arisen over the rigor and transparency of the derivation of this TTC value, including the lack of reliability of the TD50 value as a point of departure (PoD) for estimating carcinogenic potency (EFSA, 2009). As an alternative, benchmark dose modelling and the derivation of a benchmark dose level (BMD) value with a benchmark response (BMR) of 10% is recommended (e.g., Hardy et al., 2017). Benchmark dose modelling uses dose-response data to estimate the shape of the overall dose-response relationship per tissue/gender and tumor type. The BMD is a dose level, estimated from the fitted dose-response curve, associated with a specified change in response, (e.g., a BMR of 10%), whereas the benchmark dose limit (BMDL) is the corresponding lower 95% confidence bound of the BMD value (EFSA Scientific Committee et al., 2017).

In 2004, Kroes et al. showed that while the TTC concept is generally applicable, the TTC value of 0.0025 µg/kg bw/d is not appropriate for highly bioaccumulating substances and/or potent carcinogens. As a result, five chemical classes were identified and termed the Cohort of Concern (CoC): the highly potent carcinogens aflatoxin-like, azoxy, and N-nitroso compounds; polyhalogenated dibenzo-p-dioxin analogues, which bioaccumulate; and steroids. To address this for drug impurities that were members of the CoC, the European Medicines Agency (EMA), recently established a specific threshold of 18 ng/p/day for N-nitrosamines based on the most potent carcinogens in the analyzed category[2].

A critical review of the TTC approach for DNA-reactive compounds suggested several areas for improvement, relating in particular to the underlying database and the methods used to derive the thresholds (Boobis et al., 2017). The recently completed CEFIC projects LRI B18 and B18-2 addressed the following five areas:

1. *Revise and expand the content of the TTC data set derived from the Cancer Potency DataBase (CPDB):* The CPDB database was last extended 15 years ago, with the majority of entries from over three decades ago. The existing CPDB studies were reviewed, and appropriate data were included in a TTC project database along with data from new, high-quality chronic toxicity and cancer studies. New studies were retrieved from the ECHA CHEM database (DB), Cosmos DB, RepDose DB or extracted from other high-quality databases (DB). The added studies were originally published by NTP, EFSA, and JMPR (Joint FAO/WHO).
2. *Specificity of carcinogenic effect:* Studies in the CPDB without clear evidence on carcinogenic effects were excluded from the TTC data set following the criteria described by Boobis et al. (2017), "*The CPDB reports TD50 values that have been derived from either statistically significant findings in a single tissue, which should be included in the data set for relevant studies, or from findings observed in all tumor bearing animals (TBA), from more than one site, combined by NCI/NTP (MXA), or from more than one site, combined by Berkeley (MXB). Data from studies listed in the CPDB as 'TBA', 'MXA' or 'MXB' should be excluded from the data set, as the biological relevance of such grouping, comprising a range of pathologies and potential modes of action, is difficult to interpret*."
3. *Mode of action (MoA):* A classification scheme was developed to distinguish DNA-reactive from non-DNA-reactive carcinogens, mainly using available experimental data in combination with *in silico* models.
4. *Advanced dose-response modelling:* Different cancer potency reference values were compared for setting the TTC value, e.g., by replacing the TD50 value by benchmark dose (BMD) or benchmark dose limit (BMDL) values using an individual model or model averaging respectively.
5. *Impact of the CoC on overall distribution of reference values:* The TTC database contains substances that belong to the CoC. The effects of such compounds on the distribution of reference values in the TTC data set and the consequences for the threshold values to be derived were considered.

2 Aim and preparation of the workshop

The workshop discussed the approach and results of the CEFIC LRI B18 projects as well as the emerging alternatives for deriving TTC values for DNA-reactive carcinogens and evaluating the acceptability of the Cramer Class TTC values to be adequately protective for non-DNA-reactive carcinogens. Three topics were discussed at the workshop:

– Deriving PoDs for (DNA-reactive) carcinogens
– Assessing non-DNA-reactive structures
– Deriving thresholds: Risk management decisions

The virtual workshop was held in April 2021 and took place over three days. Twenty days in advance of the workshop, participants received four recorded presentations summarizing the motivation for updating the cancer database that serves as the basis for the TTC value for chemicals with structural alerts for DNA reactivity/genotoxicity, as well as an overview of the results obtained from the two CEFIC LRI B18 projects, together with an over-

[2] https://www.ema.europa.eu/documents/referral/nitrosamines-emea-h-a53-1490-assessment-report_en.pdf

view of the three discussion topics and the associated questions. The presentations were prepared by the partners of the CEFIC LRI project, namely Monika Batke, Chihae Yang, and Mark Cronin, and by Alan Boobis, who was involved in a related ILSI Europe Expert Group that preceded the work of the CEFIC LRI B18 project (Boobis et al., 2017) (see details below).

The first day was used to address questions on the project approach and data presented in the pre-recorded lectures. Days 2 and 3 were dedicated to discussion of the three topics, each topic being discussed for 80 min in three parallel break-out groups. The main findings and discussion points were reported at a plenary session to provide a first overview of areas of consensus within the break-out groups and to enable a discussion of overarching aspects. The outcome of the break-out discussions and the plenary discussion are summarized in the next section. The authors of this workshop report served on the Organizing Committee and helped in the preparation of and participated fully in the workshop. Some served as chair of a break-out group or as a rapporteur. About thirty-five TTC experts from academia, industry, and regulatory authorities participated at the virtual workshop. These experts were not involved in the CEFIC LRI projects. The following lectures were prerecorded:

– **Alan Boobis** introduced the background of the current TTC values and their regulatory use. He further laid out the current uncertainties related to the TTC values for DNA-reactive compounds as described in detail in Boobis et al. (2017). These include that the content of the CPDB is outdated, that the mode of action of the carcinogenic compounds it contains is not known, that it would be beneficial to distinguish between DNA-reactive and non-DNA-reactive carcinogens, and that other reference values such as BMD and BMDL may be more appropriate for describing carcinogenic potency.
– **Chihae Yang** gave an overview on the update of the TTC cancer database as realized in the CEFIC LRI B18 project. Among the most important criteria for the selection of the studies was that only compounds with a defined structure were considered, and compounds without clear evidence of a tumor-producing effect were excluded. With these requirements, all chronic toxicity and cancer studies were considered provided that the main study parameters were given such as species, exposure route, study duration, and more than one dose group was tested. Tumor incidences reported for mixed or summary parameters such as total incidences for adenomas and carcinomas per animal were not considered appropriate for establishing a reference value (see Boobis et al., 2017). A generic dose-response model, comparable to benchmark dose modelling, was developed to derive an effective dose (EDx) based on tumor incidence data. In addition, benchmark dose levels were obtained from the same tumor incidence data using model averaging as implemented in the Proast software[3]. TD50, EDx, and BMDL values are all reported as cancer potency reference values (PoD) for each study, data permitting.
– **Mark Cronin** gave an overview of the classification scheme differentiating between DNA-reactive and non-DNA-reactive carcinogens. The classification scheme assigns DNA reactivity to carcinogenic compounds by considering relevant *in vivo* and *in vitro* studies, followed by QSAR predictions (from commercial and non-commercial models). The impact of compounds belonging to the CoC, identified using ToxPrints in ChemoTyper[4], on the distribution of reference values for DNA-reactive compounds was determined.
– **Monika Batke** gave an overview of a recently published evaluation of non-DNA-reactive carcinogens (Batke et al., 2021). The comparison of NOAELs, ED10, and BMDL10 values for 137 non-DNA-reactive carcinogens revealed no major differences between the overall distributions and the derived 5th percentiles thereof. NOAELs were used to compare these compounds to the current Cramer Classes with and without the exclusion of bioaccumulating compounds and steroids, which belong to the CoC. The NOAEL values of non-DNA-reactive carcinogens overlapped considerably with the values of the original Munro et al. (1996) dataset in the Cramer Classes (1 and 3, there were very few compounds in class 2).

3 Topics discussed at the workshop

The following section provides an overview on the three topics and subquestions discussed in the break-out groups. Some aspects of the presentations, which are briefly outlined here again, were provided as background information to the participants. The authors have taken the freedom to prioritize and summarize the most salient points of the workshop discussion for this report.

3.1 Topic 1: Deriving PoDs for (DNA-reactive) carcinogens

All substances classified as DNA-reactive in the TTC data set are represented by at least one high-quality chronic study with increased tumor incidences. The mode of action classification was based on relevant mutagenicity information following a systematic multi-level approach, taking into account findings from *in vivo* and *in vitro* studies (mainly OECD 471 test) followed by *in silico* decisions (using several (non)commercial models). Three questions on DNA-reactive compounds were discussed:

Question 1: Is it most appropriate to use all carcinogens or to only use the PoDs from DNA-reactive carcinogens?

The original cancer TTC thresholds were based on distributions of data for all carcinogens, regardless of their mode of action. The CEFIC LRI project team proposes to use only the DNA-reactive carcinogens to assess the threshold for DNA-reactive structures, as these carcinogens are on average more potent compared to non-DNA-reactive compounds. Exclusion of the latter therefore avoids a potential dilution. In the tiered de-

3 https://www.rivm.nl/en/proast

4 https://chemotyper.org

cision tree TTC concept, the non-DNA-reactive structures are assessed according to their Cramer Classes or category-specific thresholds such as the one for organophosphates/carbamates.

Report from break-out groups

The classification "DNA-reactive" raised several questions for clarification, e.g., based on which data and by which approach were DNA-reactive compounds classified, to which extent are clastogenic compounds included, or is clastogenicity regarded as an additional mode of action? This is important because clastogenicity is generally considered to have a threshold such that linear extrapolation may not be appropriate (although it was noted that the decision on whether to apply the TTC for genotoxic carcinogens to such a compound was distinct from whether to include such compounds in the derivation of the TTC values). It was stated that the focus of this mode of action classification must be on DNA-reactive mutagens, for which TTC is usually applied. In the LRI B18 project, the classification was formalized into a decision tree, whereby preference was given to experimental data relating to DNA reactivity (e.g., Ames test results) and, should insufficient data be available, subsequently structural alerts and QSAR predictions for Ames tests were applied.

A consensus was reached regarding the restriction of the TTC data set to DNA-reactive compounds. Separating DNA-reactive compounds was seen as appropriate as the resulting TTC values will also only be applied to such compounds based on the presence of structural alerts. It was however recommended to show the impact of the in- and exclusion of non-DNA-reactive compounds on the overall PoD distribution, and the derived thresholds thereof, to provide a robust scientific rationale and a fully transparent approach.

The application of a TTC value to DNA-reactive compounds requires a robust and scientifically sound classification approach, also considering that for most untested compounds only structure and physicochemical properties will be available. Open questions remain with regard to the reproducibility of the classification (e.g., several Ames tests available) and the type of evidence considered appropriate (bacterial mutagenicity versus clastogenicity versus *in silico* predictions).

It was noted that QSAR predictions are often very conservative. Rules on how to use and combine results from *in silico* tools must be defined. To date the ICHM7 guideline[1] recommends using at least two different QSAR models, a knowledge-based and a statistical approach, followed by an expert review to conclude on DNA reactivity. In this context, a need for structural alerts based on tertiary structure (3D) was expressed.

Question 2: Is it appropriate to assess the TTC threshold for DNA-reactive structures excluding structures of the exclusion categories?

Currently, the TTC data set includes compounds belonging to the CoC, and it has been shown that the TTC value of 0.15 µg/p/d is not sufficiently protective for such highly potent carcinogens (Kroes et al., 2004). Substances belonging to the CoC can therefore not be evaluated using the current TTC value.

The project team presented analyses on the PoD distributions in the updated TTC data set with and without consideration of the structures belonging to the CoC. These substances accumulate at the low end of the distribution. Where a low percentile such as a 5th percentile of a BMD(L) distribution is used as the starting point to derive a TTC value, the inclusion of highly potent carcinogens, outside the applicability domain of the TTC concept, will lead to very low threshold values.

Report from break-out groups

From a scientific point of view, the TTC data set and the corresponding TTC value should be restricted to the compounds in their applicability domain. The view was expressed that the decision on their in- or exclusion is not a scientific but a policy question. The derivation of TTC values based on a 5th percentile approach was discussed, e.g., starting from a BMD or BMDL distribution. When starting from a 5th percentile, a concern expressed at the workshop was that keeping the CoC in the data set could result in extremely low and thus potentially overprotective thresholds for regulating DNA-reactive compounds. The applicability and usefulness of such TTC values have further been questioned as levels may be too low to be routinely measured and quantified in regular product release analyses.

It was recommended to quantify the impact of the CoC compounds on the resulting PoD distribution, e.g., by performing a sensitivity analysis on covariates.

Other exclusion categories were mentioned, e.g., metals, polymers, proteins, radiolabeled compounds, biologics. These compound classes are not well represented in the CPDB, and thus an analysis of their impact is not feasible. It was noted that no new categories for CoC were identified in the last decades (EFSA and WHO, 2016). However, Cross and Ponting (2021) recently published a subcategorization of N-nitrosamines, which differentiates less from highly potent nitrosamines based on the modifying impact of substituents, e.g., bulky side chains or electron-withdrawing groups. The present data set might offer the opportunity to reevaluate the current CoC classes and potentially identify new classes/compounds.

Question 3: Is there a strong preference for using concentration-response or benchmark dose data?

The cancer TTC proposed by Kroes et al. (2004) was derived from the FDA's Threshold of Regulation and was based on TD50 distributions. It was later questioned whether this approach remains a sound methodology for the evaluation of carcinogenic compounds. The project team presented analyses on the DNA-reactive data set for PODs derived from concentration-response (10% and 50% tumorigenic responses) and from benchmark dose calculations. There is varying coverage of the differently derived PoDs due to the differences in data requirements for concentration-response in comparison to benchmark dose methods. PoDs, e.g., BMD10 calculated from benchmark doses, are lower than any corresponding concentration-response (TD50) values, and this may be of concern with regard to future lowering of TTC values. The BMD calculations provide confi-

dence intervals, which may have advantages, e.g., narrower confidence intervals indicate less uncertainty in the data.

Report from break-out groups
A number of issues were discussed, and it was not possible to reach an overall conclusion.

The ED10 value is a central point estimate, and the shape of the dose response curve can be directly compared. A BMDL10 is the lower 95th percentile confidence interval of the benchmark dose value. It therefore includes its uncertainty and is used for chemical-specific risk assessment. The discussion centered on the question of whether the TTC values should be based on the confidence interval.

A BMD value, in addition to the corresponding BMDL, was considered to be informative, and the use of BMD software other than PROAST was recommended as PROAST does not provide an estimate of the BMD values. It was noted that a central estimate such as BMD/ED10 is less variable than the associated confidence interval when comparing different dose-response models. Since a 5th, 10th or 50th percentile of a global distribution is used to derive the TTC value, it was argued that a central estimate such as the ED10 was the better choice, as otherwise conservative assumptions could accumulate.

Another suggested approach is to use the BMD in combination with a descriptor of variance calculated based on the BMDL/BMDU (lower and upper limit) values to characterize uncertainty. Other experts preferred the BMDL10 value, as it accounts better for the uncertainty in the BMD value but noted that it would be good to have a central estimate to compare with.

The project used the lowest derived reference dose per compound (min value approach) for the distribution analyses; other strategies such as hierarchical approaches were proposed.

3.2 Topic 2: Assessing non-DNA-reactive structures

Question 1: Is the exclusion of compounds belonging to the exclusion categories appropriate?

The project team decided to remove substances with steroid structures and strongly bioaccumulating structures (such as dioxins and dioxin-like PCBs) from the TTC data set of non-DNA-reactive carcinogens. Both compound classes belong to the exclusion categories, which are generally accepted and part of many regulatory applications (EFSA, 2012; EFSA Scientific Committee et al., 2019). While strongly bioaccumulating and steroidal compounds were observed frequently at the lower end of the PoD distribution in this data set, no other specific compound class accumulated there. Both groups are highly toxic, and their exclusion increased the 5th percentile more than when randomly excluding the same number of other compounds from the data sets. Should additional structures be considered for exclusion?

Report from break-out groups
Arguments were raised supporting the exclusion of strongly bioaccumulating and steroidal compounds. The dosimetry is very different for strongly bioaccumulating compounds, and therefore these substances should be excluded from the TTC data set. The decision criteria have to be reported in a transparent and understandable way. Steroids are of concern because of their endocrine activity, and they may differ in their potency with regard to, e.g., ER binding. It was questioned whether such data could be included as one decision factor into the analysis. This aspect can probably not be taken into account, as these compounds are currently classified based only on shared and typical structural properties and not on data characterizing their biological mode of action. Generally, there was an agreement with the exclusion of the classes reported here, and the question was raised whether more groups need to be analyzed, e.g., endocrine-disrupting compounds that do not belong to the class of steroids but are still carcinogens, or PFAS (perfluoroalkyl and polyfluoroalkyl substances) as bioaccumulating substances. A sensitivity analysis was recommended to evaluate the impact of the in- and exclusion of different classes. The analysis can then be used to show in a systematic and transparent way why it is important to exclude such compounds from the current TTC concept. These evaluations should also focus on species differences, e.g., in bioaccumulation or metabolism of chemicals, to elucidate the extrapolation from rodent to human in the context of TTC.

From an implementation point of view, it was noted that while it is easy to identify steroids, the identification of potentially bioaccumulating compounds is more challenging. This is a general unsolved problem in the TTC approach as bioaccumulating compounds are already part of the exclusion criteria.

Question 2: Which PoD is appropriate to evaluate TTC values for non-DNA-reactive compounds?

The data set comprises different reference values (PoDs) for non-DNA-reactive carcinogens, in particular no observed effect levels (NOELs), BMDL10, and ETD10 values per compound.

NOEL values covered compounds and their studies to a large extent, as all studies in the data set identified a NOEL, while modelling of ETD10 and BMDL was not possible for all studies. The obtained distributions of BMDLs and NOELs are comparable; ETD10 values showed a slight but statistically non-significant shift to higher values. Is the selection of NOEL values appropriate?

Report from break-out groups
The discussion started with some clarifications. The LRI B18_2 project extracted NOEL values from peer-reviewed publications, the uncertainty and comparability of values is therefore not known. Only studies of high quality, comprising a reasonable dose range, were used. BMD(L) derivation was not always possible based on the reported tumor incidences and number of tested doses. The main difference between the BMD(L)and NOEL values is that the BMD(L) values are derived from tumor incidences, whereas the NOEL is determined from non-neoplastic as well as neoplastic lesions. A tumor occurred at the LOEL in 50% of the studies.

It was discussed that the endpoint carcinogenicity is generally not the critical concern for non-DNA-reactive compounds because exposure limits established for noncancer endpoints will also be protective for cancer by a non-DNA-reactive mode of action; this might contribute to the finding that BMD(L) values are not markedly different from NOELs. Moreover, it was stated that the difference between BMD and NOEL values should not be high in high-quality studies. BMDL10 and ED10 values were derived from the same tumor incidence data and their dose response data. However, the two modelling approaches do show some differences with regard to obtained values and data sets that could not be modelled. An investigation of these differences is needed to better understand the robustness and reliability of both approaches.

It was, however, agreed that a benchmark model approach (by either model) is more precise compared to a NOEL, which depends on dose spacing and dose selection. The preferred value is therefore in principle a BMD or BMDL for the assessment of carcinogenic potency. However, with regard to the TTC analyses, it was agreed that a NOEL is a valid reference value and is adequate for the analysis of threshold values. The analysis is restricted to the available data, and in this data set, a NOEL was available for more studies and therefore covers more compounds compared to the BMD(L)10 data set. This finding can be explained by taking into account that the studies in the TTC dataset were not conducted with the aim of facilitating BMD modelling and therefore often only 2 to 3 dose groups were tested. The objective of the present study is to compare the PoDs of the DNA-reactive substances with the existing TTC values of the Cramer Classes. So far, there is no consensus on the extrapolation step of BMD(L) value, and the implications of using a different PoD remain unknown. The use of NOELs was therefore seen to be appropriate to compare to the current Cramer Class thresholds.

Question 3: Are TTC values from Cramer Classes applicable to non-DNA-reactive carcinogens?

It is possible to identify structural alerts for DNA reactivity for a chemical lacking toxicity data, but in the absence of such alerts it would not be known (nor is it easily predicted) whether the chemical would test positive in a rodent bioassay by a non-genotoxic mode of action. While the use of the tiered TTC assumes that the Cramer Class TTC levels are adequately protective for non-genotoxic carcinogens, a confirmatory analysis had not previously been conducted.

The comparison of the NOELs in the TTC data set for non-DNA-reactive carcinogens (Batke et al., 2021) showed a good overlap with the values of the Cramer Classes originally derived by Munro et al. (1996). Most of the non-DNA-reactive substances of the project data set belong to Cramer Class III, some to Cramer Class I. Because of the paucity of data in Cramer Class I, no further comparative analyses were made, as their informative value is limited. The 5th percentiles of Cramer Class III non-DNA-reactive carcinogens and the original Munro data set are comparable. For Cramer Class III, a random leave-out of 5% of substances in both Munro and non-DNA-reactive cancer data sets results in similar ranges of TTC values, indicating the robustness of the values. This led the project team to the conclusion that the current Cramer Classes are adequately protective for chemicals lacking an alert for DNA reactivity, regardless of whether they might ultimately test positive in a rodent bioassay, when excluding the CoC.

Report from break-out groups

The conclusion that Cramer Classes are protective for non-DNA-reactive carcinogens is generally supported. Nevertheless, it was recommended that larger data sets becoming available in future should be analyzed in more detail. A traditional statistical test like the non-parametric Kolmogorov-Smirnov test might be useful to compare different distributions. If more Cramer Class I compounds become available, an enhanced assessment would also be desirable for this class of non-DNA-reactive chemicals. Finally, the integration of BMDL values would be desirable in future assessments.

It would be highly interesting to learn more about the different mechanisms leading to tumor formation of non-DNA-reactive compounds, as the cancer endpoint most likely progresses from other adverse effects within the course of the chronic or cancer study, e.g., through immunosuppression, cytotoxicity, oxidative stress, or chronic inflammatory processes. However, some panelists suggested that this will not have any bearing on the TTC.

Another interesting exercise would be to set the obtained results into context, e.g., by comparing the obtained results with the six TTC classes proposed by FDA[5]. The updated data sets might also be a good starting point to derive category-specific threshold values.

3.3 Topic 3: Deriving thresholds: Risk management decisions

The discussion of topic 1 and 2 focused on the choice of the appropriate data set and setting of in- and exclusion criteria as well as of appropriate reference values to characterize the carcinogenic potency of substances. The next step is the derivation of TTC thresholds, in which different approaches can be used for extrapolation and assumptions have to be made regarding the level of protection.

Question 1: Is the objective to confirm the adequacy of the current threshold to protect human health or is it to derive new thresholds based on the analysis?

The workshop participants agreed that both options are valid and that one does not exclude the other. It was emphasized that the curated data set has a high value in itself, as it can be used for category and subgrouping approaches in addition to the TTC analyses. It was recommended to publish the entire curated data set as well as the methods and models used in order to increase transparency, reproducibility of the values, and understanding of the methods.

5 https://www.fda.gov/media/144891/download

A consensus was achieved that the new data should first be used to determine whether the existing TTC values can be supported. However, it was also noted that the established threshold of 0.15 µg/p/day has been reviewed several times and no major concerns have been associated with it to date, so a confirmation of the adequacy of the current threshold is not a priority.

The established threshold of 0.15 µg/p/day first published by Kroes et al. (2004) has its origins in the U.S. FDA's original Threshold of Regulation (ToR, 1.5 µg/p/day) (Cheeseman et al., 1999). The ToR is derived from the TD50 value distribution but is not based on a specific percentile of this distribution. In the context of the re-evaluation of the TTC data set, this data gap could be closed in order to achieve a more transparent decision and thus a more precise scientific basis. The selection of the percentile and the extrapolation approach should be described in detail. In addition, the reassessment allows the threshold to be based on updated reference values, such as BMDL(10). The analyses presented at the workshop support the current thresholds, and it was emphasized that differences between new and original values are within uncertainty of the methods (Batke et al. 2021; publication on LRI B18 in preparation).

A longer-term goal could be to derive new threshold values based on the present analysis and to start the process of regulatory acceptance by providing data on reliability and robustness of the approach. Approaches for extrapolating BMD/BMDL values to determine a human threshold need to be developed and fine-tuned, as use of TD50 values is considered to be outdated. Beside thresholds for global classes of DNA-reactive or non-reactive compounds, the data set might be explored to develop potency-based subclasses for DNA-reactive compounds. If the chemical space can be extended, different classes of DNA-reactive compounds might be identifiable, e.g., ranging from high to low potency. This could significantly increase the utility of TTC as the potency for carcinogens ranges over at least 6 orders of magnitude. The use of new approach methodologies within integrated approaches to testing and assessment might be an option to further expand the data set and to integrate mechanistic information.

Question 2: Which combination of reference value type and percentile is preferable as starting point for risk extrapolation?

To answer this question, several analyses were recommended to evaluate the most appropriate approach and to learn more about data variability. For this purpose, all values, BMDL10 as well as ED10 values, should be used. From the distribution of one value or the other, the workshop participants favored a percentile like the 5th percentile as the starting point for threshold derivation, as it takes into account a major part of the distribution while being more robust to outliers than smaller percentiles.

Another discussion addressed the different protection goals from BMDL to BMD. To date, TTC for genotoxic compounds is based on a linear extrapolation of TD50 values to calculate the exposure associated with a 1/10^6 risk. For BMD or BMDL values, alternative approaches need to be developed and agreed. It is noted that a margin of exposure of 10^4 when using a BMDL10 is equivalent to a risk of 1/10^5 if a linear relationship is assumed, which is a protection goal used by many regulators. In the end, this remains a policy decision.

Finally, it was recommended to consider how the most appropriate approach relates to a protective dose and to take into account potential differences between US and European regulatory methodology.

Question 3a: Which percentile is appropriate to conclude on TTC values for DNA-reactive compounds?

The choice of the percentile was considered to depend on the representativeness of the data set for the chemicals that we are concerned about. This comment raised the question of whether a database with several hundred substances is representative of "the world of chemicals", pointing out that this is theoretically not possible. However, re-evaluations of the cancer database over time showed little impact on the distribution of potencies. As described by Cheeseman et al. (1999), FDA's Threshold of Regulation was originally based on an analysis of 477 carcinogens; expansion of that data set to include 709 carcinogens did not have a significant impact on the range of potencies. Kroes et al. (2004) expanded the data set to 730 compounds, again confirming no significant change to the potency distribution. Relatively few carcinogens have been added to the data set since then. For this reason, the current database was considered sufficiently representative of the "world of carcinogens known to date". It was noted that epidemiological studies have not revealed any new carcinogens.

In the following, the use of a low percentile, such as the 5th percentile, as starting point for the derivation of TTC values for DNA-reactive carcinogens was discussed. As expected, the project analyses show that highly potent carcinogens such as the CoC occur mainly at or below the 5th percentile in the data set. The CoC will have to be excluded from the distribution used to establish the TTC values, otherwise they will lead to very low values, protective of compounds outside the applicability domain of the TTC concept.

Lower percentiles (below the 5th) would be based on only very few compounds, and this raises the question of the robustness of the value and its applicability for regulatory purposes. The analyses presented at the workshop illustrating options to derive TTC values based on BMD(L) values indicated that the current values are confirmed by the range of "new" values at a screening level of 1/10^5 and 1/10^6. In the long term, alternative approaches such as an "internal TTC", i.e., based on the absorbed systemically available dose, should be considered, which could take interspecies differences in kinetics and MoA into account (Ellison et al., 2020).

Question 3b: Should different risk levels be defined for, e.g., contaminants versus intentionally added substances and drugs?

Workshop participants expressed a consensus about the fact that human health needs to be protected regardless of the pur-

pose of the exposed compound. There is no scientific reason for defining different thresholds for safe levels of unintended exposure to compounds, e.g., contaminants in food versus intentionally added substances such as drugs. However, other considerations such as risk/benefit and differences in exposure pattern, such as lifetime versus short-term exposure, may need to be considered.

Question 3c: Which safety factors are needed to assure no impact on human health? Which acceptable risk level is generally appropriate?

All working groups agreed that a safety factor must be defined by risk managers, as this is not a scientific but a policy issue. This question cannot, therefore, be answered by updating and re-analyzing the TTC data set or within this workshop. Risk management decisions might differ for drugs because of risk/benefit considerations compared to DNA-reactive chemicals to which humans are exposed at home or at the workplace. Acceptable risk levels may differ for occupational versus consumer exposure or potentially susceptible human populations such as infants, pregnant women, or elderly people, etc.

The different needs are currently reflected in the $1/10^5$ risk for drugs versus $1/10^6$ for chemicals occurring in cosmetics or food. The choice of acceptable risk level also varies with different regulations (e.g., California Proposition 65 has No Significant Risk Levels based on a risk level of $1/10^5$).

The project can contribute to a better understanding of these issues by comparing various approaches such as linear extrapolation with other models and by addressing the uncertainty of the different approaches. A comprehensive view was considered to be helpful, e.g., by reporting all potential thresholds based on most common risk values ($1/10^5$; $1/10^6$; worker safety $4/10^5$; others).

From a user's perspective, it is of utmost importance that the final approach is transparent and clear in order to gain confidence in the derived threshold value.

4 Summary and conclusions

The workshop concluded that the updated and curated database improved substantially on the existing CPDB with regard to data quality, study selection, and derivation of modern reference values. The focus on carcinogens avoids the derived threshold being "diluted" by irrelevant data. Categorization according to mode of action was seen as an added value as well as resolving the question of the contribution of non-DNA-reactive compounds.

Sustainability is one open issue for the TTC database, as a constant update with newly published high-quality data would be an advantage. The release of the database to the public was strongly recommended by the workshop participants. They also pointed out that search and analysis functions, such as on structural similarity in the context of grouping approaches, would be of great benefit.

The data set was found to contain several sources of uncertainty, e.g., regarding chemical selection, data variability of the BMD(L) values, etc. In this context, statistical analyses like bootstrapping were proposed to derive a good estimate of the robustness of, e.g., the 5th percent level. It was also discussed whether hierarchical or weight-of-evidence strategies could be used as an alternative to the min value approach to select the most scientifically defensible value instead of the lowest. Such analyses need to be further defined in terms of feasibility and relevance.

Depending on the selection of the PoD, preferably BMD(L) values, a new assessment concept for the derivation of threshold values has to be developed and will need assessment factors other than $1/10^6$ to derive a threshold. The final choice of an assessment factor was seen as a policy rather than a scientific decision. Nevertheless, the analysis of different options and the illustration of residual uncertainties and the robustness of such thresholds could help to increase confidence in the derived values/approaches.

The updated database will allow exploration of some subclasses for which specific thresholds can be derived comparable to, e.g., the newly developed thresholds for nitrosamines. Machine learning approaches may be helpful to cluster compounds according to their mechanistic features and their observed potency into categories or broader compound classes comparable to the Cramer Classes for non-DNA-reactive compounds.

A recommendation on which approaches and tools should be used to make the call "DNA-reactive" would be a very valuable outcome of the project, outlining a "best set of rules". The classification of compounds as non-DNA-reactive carcinogens was also seen as desirable, although this is currently not possible based on structural properties. One perspective could be to consider the risks of other mode of action groups such as genotoxicity, endocrine disruption, direct or cytotoxic mitogens.

Future work could aim at a better understanding of properties/mechanisms leading to differences in potency of DNA-reactive compound classes, e.g., by considering differences in toxicokinetic (absorption, metabolism, distribution, and excretion (ADME)) as well as in DNA adduct and repair processes. This could lead to a different basis for establishing acceptable thresholds.

References

Batke, M., Afrapoli, F. M., Kellner, R. et al. (2021). Threshold of toxicological concern – An update for non-genotoxic carcinogens. *Front Toxicol 3*, 688321. doi:10.3389/ftox.2021.688321

Boobis, A., Brown, P., Cronin, M. T. D. et al. (2017). Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their re-evaluation. *Crit Rev Toxicol 47*, 705-727. doi:10.1080/10408444.2017.1318822

Cheeseman, M., Machuga, E. and Bailey, A. (1999). A tiered approach to threshold of regulation. *Food Chem Toxicol 37*, 387-412. doi:10.1016/s0278-6915(99)00024-1

Cramer, G. M., Ford, R. A. and Hall, R. L. (1976). Estimation of toxic hazard – A decision tree approach. *Food Cosmet Toxicol 16*, 255-276. doi:10.1016/S0015-6264(76)80522-6

Cross, K. P. and Ponting, D. J. (2021). Developing structure-activ-

ity relationships for N-nitrosamine activity. *Comput Toxicol 20*, 100186. doi:10.1016/j.comtox.2021.100186

EFSA (2009). Guidance of the Scientific Committee on Use of the benchmark dose approach in risk assessment. *EFSA J 7*, 1150. doi:10.2903/j.efsa.2009.1150

EFSA (2012). Scientific opinion on exploring options for providing advice about possible human health risks based on the concept of threshold of toxicological concern (TTC). *EFSA J 10*, 2750. doi:10.2903/j.efsa.2012.2750

EFSA Scientific Committee, Hardy, A., Benford, D. et al. (2017). Update: Guidance on the use of the benchmark dose approach in risk assessment. *EFSA J 15*, 4658. doi:10.2903/j.efsa.2017.4658

EFSA Scientific Committee, More, S. J., Bampidis, V. et al. (2019). Guidance on the use of the threshold of toxicological concern approach in food safety assessment. *EFSA J 17*, e05708. doi:10.2903/j.efsa.2019.5708

EFSA and WHO (2016). Review of the threshold of toxicological concern (TTC) approach and development of new TTC decision tree. EFSA supporting publication. doi:10.2903/sp.efsa.2016.EN-1006

Ellison, C. A., Api, A. M., Becker, R. A. et al. (2020). Internal threshold of toxicological concern (iTTC): Where we are today and what is possible in the near future. *Front Toxicol 2*, 621541. doi:10.3389/ftox.2020.621541

Hardy, A., Benford, D., Halldorsson, T. et al. (2017). Update: Use of the benchmark dose approach in risk assessment. *EFSA J 15*, e04658. doi:10.2903/j.efsa.2017.4658

Kobets, T. and Williams, G. M. (2019). Review of the evidence for thresholds for DNA-reactive and epigenetic experimental chemical carcinogens. *Chem Biol Interact 301*, 88-111. doi:10.1016/j.cbi.2018.11.011

Kroes, R., Renwick, A. G., Cheeseman, M. et al. (2004). Structure-based thresholds of toxicological concern (TTC): Guidance for application to substances present at low levels in the diet. *Food Chem Toxicol 42*, 65-83. doi:10.1016/j.fct.2003.08.006

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

Preston, R. J. and Williams, G. M. (2005). DNA-reactive carcinogens: Mode of action and human cancer hazard. *Crit Rev Toxicol 35*, 673-683. doi:10.1080/10408440591007278

Russell, W. M. S. (1999). *The Principles of Humane Experimental Technique*. Special edition. Wheathampstead: Universities Federation for Animal Welfare. ISBN0900767782.

WHO (2020). *Principles and Methods for the Risk Assessment of Chemicals in Food* (section 4.5). 2nd edition. Geneva, Switzerland: World Health Organization. ISBN9789241572408.

Conflict of interest

The authors state that they have no conflict of interest.

Acknowledgements

The authors are grateful to the participants of the workshop for their active participation in the discussion. This work was supported by the ILSI and CEFIC LRI.

Sylvia E. Escher[1], Susan P. Felter[2], Heli Hollnagel[3], Alan R. Boobis[4], Chihae Yang[5], James Rathman[5], Mark T. D. Cronin[6] and Monika Batke[1]

[1]Fraunhofer ITEM, Hannover, Germany; [2]Procter & Gamble, Central Product Safety, Mason, OH, USA; [3]Dow Europe GmbH, Horgen, Switzerland; [4]National Heart and Lung Institute, Imperial College London, London, UK; [5]Molecular Networks-Altamira, Nürnberg, Germany; [6]Liverpool John Moores University, Liverpool, UK

(sylvia.escher@item.fraunhofer.de)