Archives of Toxicology
https://doi.org/10.1007/s00204-024-03951-8

REGULATORY TOXICOLOGY

Critical comparison of BMD and TD_{50} methods for the calculation of acceptable intakes for N-nitroso compounds

Robert Thomas[1] · David J. Ponting[1] · Andrew Thresher[1] · Joerg Schlingemann[2] · John W. Wills[3] · George E. Johnson[4]

Received: 16 September 2024 / Accepted: 17 December 2024

Abstract
The tumorigenic dose 50 (TD_{50}) is a widely used measure of carcinogenic potency which has historically been used to determine acceptable intake limits for carcinogenic compounds. Although broadly used, the TD_{50} model was not designed to account for important biological factors such as DNA repair and cell compensatory mechanisms, changes in absorption, etc., leading to the development of benchmark dose (BMD) approaches, which use more flexible dose–response models that are better able to account for these processes. Using a nitrosamine dataset as a case study, we compare the impact of moving to a BMD-based limit as opposed to a TD_{50}-based limit. Although there are differences in individual potency estimates between the two approaches for some compounds, we show that the key metrics such as the 5th percentile of the respective potency distributions, used when calculating class-specific default acceptable intakes, are not greatly affected. Furthermore, potency estimates for nitrosamine compounds relevant to read-across do not vary by more than a factor of 3, which is little in the context of the inherent variability in a biological response, in an overall landscape wherein potencies can vary by four orders of magnitude. This suggests a move to BMD-based limits is achievable without significant disruption to existing limits while utilising a more robust methodology.

Keywords Nitrosamines · Acceptable intake · TD_{50} · Benchmark dose (BMD) · Dose–response modeling · Drug impurities

Introduction

The assessment of a compound's carcinogenic potency commonly requires an extrapolation of data from high-dose animal studies to much lower levels that are relevant for human exposure. In other words, to evaluate the cancer risk of a specific chemical for humans, one needs to connect the exposure level at which substantial tumor incidence is observed in animal studies and the level to which humans can realistically be exposed or the level at which the residual cancer risk is deemed acceptable. This extrapolation requires two things: a starting point (point-of-departure) and a mathematical model to be applied for the extrapolation. Applicable models were previously discussed and compared elsewhere (Edler et al. 2002). For the assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals, linear extrapolation is the default approach defined in guideline ICH M7 (ICH 2017), and it is also preferred by the European Food Safety Authority (EFSA) for the risk assessment of carcinogenic substances in food (European Food Safety Authority 2005). In the presented work, we compare two mathematical models that derive points of departure from animal-study dose–response data enabling subsequent linear extrapolation.

The determination of the acceptable intakes for both compounds with or without carcinogenicity data [for which a threshold of toxicological concern, TTC, has been determined in the ICH M7 guidance, based on evaluation of available TD_{50} values (Kroes et al. 2000, 2005, 2004; Müller

✉ Joerg Schlingemann
joerg.schlingemann@merckgroup.com

✉ George E. Johnson
g.johnson@swan.ac.uk

1 Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, UK

2 Merck KGaA, Frankfurter Str. 250, 64293 Darmstadt, Germany

3 GlaxoSmithKline R&D, Gunnels Wood Road, Stevenage, UK

4 Institute of Life Science, Swansea University Medical School, Swansea, UK

Published online: 03 January 2025

et al. 2006; Munro 1990)] is critically dependent on these methods of potency quantification. Recent concerns as a result of nitrosamine impurities in marketed pharmaceuticals have led to a particular focus on these values and methods used to determine them (Thomas et al. 2021) since control of nitrosamine impurities to the TTC is not considered sufficiently protective of carcinogenic risk as they fall under the so-called 'cohort of concern'. This requires evaluation of compound-specific data to determine limits that can be exceptionally low—at the boundaries of analytical feasibility both for detection and control (Burns et al. 2023)—and the precise values of the limits therefore derived can easily make the difference between a drug being available to market and not.

The TD_{50} model

The TD_{50} method uses an exponential model of tumour incidence. It makes the simple assumption that a given increase in the dosage of a compound will cause tumours in a set fraction of animals exposed. The TD_{50} is then the dose that halves the probability of remaining tumour free. As an example, if you had 80 animals exposed to the compound at the TD_{50} and there was a 10% chance of developing a tumour in the control group, you would expect 40 of them to develop tumours due to the compound, another 4 by chance (10% of 80–40), and 36 (45%) to remain tumour free. In turn, if exposed to a dose double the TD_{50} you would expect half of the tumour-free animals to then develop tumours leaving 18 (22.5%) animals tumour free, and if exposed to three times the TD_{50} you would halve the remaining tumour-free animals again leaving 9 (11.25%) animals tumour-free. This differs from a linear model which would predict no tumour-free animals at double the TD_{50} and cannot give sensible predictions at three times the TD_{50}. The TD_{50} model's dose–response relationship [derived from Peto et al. (1984)] is given by Eq. 1:

$$P(\text{tumour}) = 1 - b \bullet e^{-k \bullet \text{dose}} \tag{1}$$

b is related to the base rate of the tumours. At dose = 0 then $e^{-k \bullet \text{dose}} = 1$, so the probability of getting a tumour is $1 - b$. In other words, b is the probability of *not* getting a tumour if no drug is given. k is related to the potency of the compound and is used to calculate the TD_{50}, as shown in Eq. 2. The TD_{50} is the point at which the tumor-free rate is halved relative to the baseline b which corresponds to a tumor rate given by Eq. 3.

$$e^{-k \bullet \text{TD}_{50}} = 0.5 \rightarrow -k \bullet \text{TD}_{50} = \ln(0.5) \rightarrow \text{TD}_{50} = \frac{\ln(0.5)}{-k} = \frac{\ln(2)}{k} \tag{2}$$

$$P\left(\text{tumour@TD}_{50}\right) = 1 - b \cdot e^{-k \cdot \text{TD}_{50}} = 1 - b \times 0.5 = 1 - \frac{b}{2} \tag{3}$$

As demonstrated in Fig. 1, and against common belief, linear back-extrapolation from the TD_{50} to a dose associated with a lower tumour incidence is not conservative compared with the exponential function: i.e., it slightly overestimates the dose associated with a particular tumour incidence, or in other words, it underestimates the compound's potency. Please refer to Supplementary Material 1 for a calculation of the extent of the underestimation associated with linear

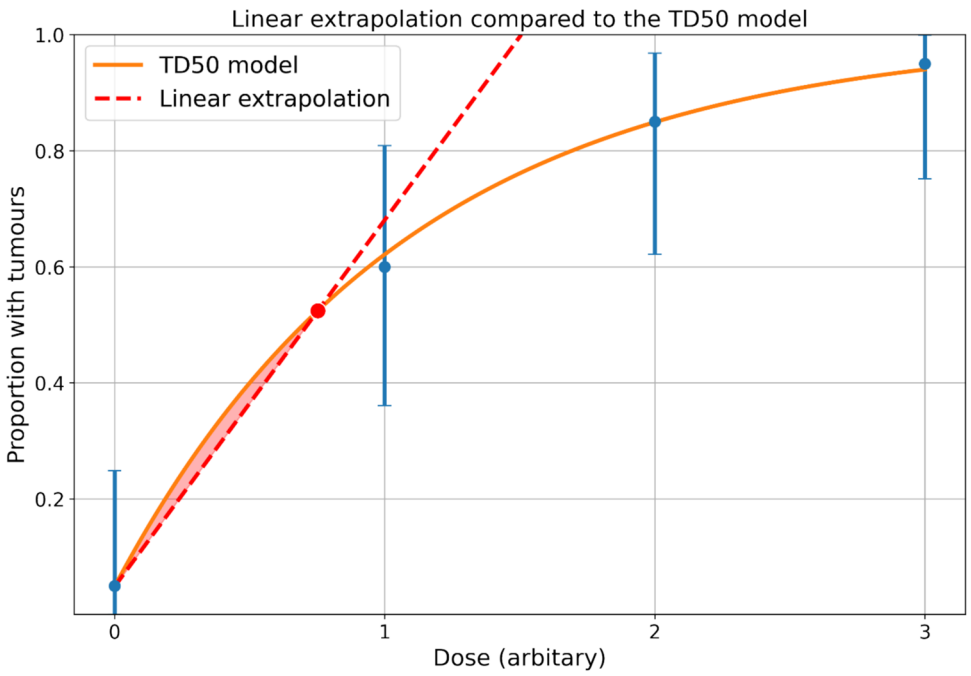

Fig. 1 Linear back-extrapolation from the TD50 is not necessarily conservative. Theoretical example of a dose–response plot (blue dots and error bars), exponential regression using the TD_{50} model (orange curve) and linear back-extrapolation (dashed red line) from the TD_{50} (red dot) to a dose with a lower tumour incidence

back-extrapolation from the TD_{50} to a tumour incidence of 1 in 100,000 (10^{-5}).

Benchmark dose approach

In the BMD approach, the model to be fitted to the data is not fixed. Rather, different candidate models are fitted to the data and compared using a goodness-of-fit criterion, and any models showing sufficiently good fits are selected for model averaging. This allows a more accurate reproduction of the frequently sigmoidal dose–response curve, particularly at the lower end of the dose range. The quantitative potency estimate derived from carcinogenicity benchmark dose modelling is typically the BMD_{10}, the dose at which tumour growth is expected to increase to 10 percentage points over background level in the dosed animals (similar to a TD_{10}). Oftentimes the $BMDL_{10}$ is used as the point-of-departure estimate which is the dose representing the lower 95% one-sided confidence limit of the BMD_{10}.

As shown in Fig. 2, for a dose–response following an exponential response, the extrapolation based on the TD_{50} model and a benchmark dose model are quite similar. However, for a dose–response following a sigmoidal characteristic, which is very common for biological systems e.g. when compensatory mechanisms aren't saturated in the low-dose region, the benchmark dose model creates a much better fit and slightly overestimates the potency based on linear back-extrapolation from the $BMDL_{10}$ to a dose with lower tumour incidence. Meanwhile, back-extrapolation from the TD_{50} to such a dose would, in this theoretical sigmoidal case, result in a severe overestimation of potency. In this example, a Weibull curve has been used to represent a benchmark dose model, due to its flexibility and inclusion in many benchmark dose packages. For simplicity and ease of visualization model averaging has been waived in this example.

Here we calculate BMD and TD_{50} potency estimates for a large nitrosamine dataset and derive acceptable intakes based on both methods. These values are compared with a specific focus on those compounds relevant for read-across and in terms of the impact on the class-specific default acceptable intakes.

Methods

Data collection

Dose–response data for 55 nitrosamine compounds was taken from the Lhasa Carcinogenicity Database (LCDB) (Lhasa Limited 2023) along with compound-level summary TD_{50}s. These are the individual dose–response TD_{50} values for the most sensitive tissues from the different studies, aggregated (within species groupings, e.g., rat, mouse and monkey) by harmonic mean. The lowest value of these in a rodent species—rat, hamster, or mouse—is then selected to provide a single value per compound. The use of just rodent species is to enable better comparison across studies, whilst avoiding comparing rodents with primates or dogs, for which less data is available and the number of animals

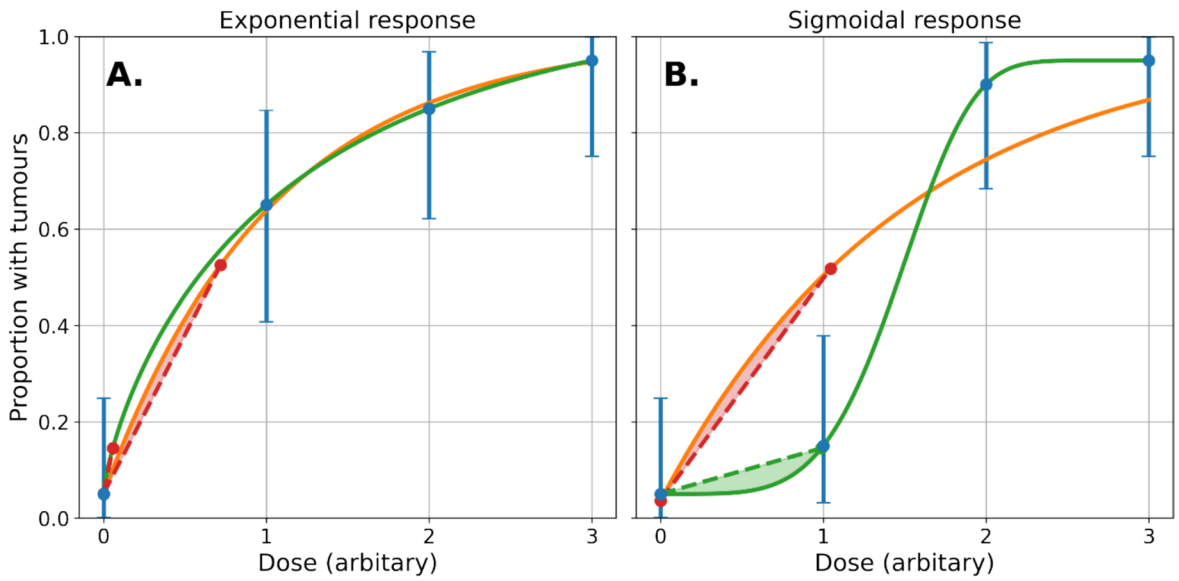

Fig. 2 Comparison of the TD_{50} model (exponential, orange curves) and a benchmark dose model (Weibull, green curves) for the extrapolation of theoretical dose responses that follow **A** an exponential and **B** a sigmoidal characteristic. Under-estimation of potency relative to the model due to extrapolation is represented by the red-shaded region, while over-estimation is shown by the green-shaded region

per dose is very low. In total, 1175 dose–response relationships were extracted.

Calculation of benchmark doses.

Benchmark doses were calculated using PROAST version 70.3. Although other benchmark dose software is available, PROAST uses a widely accepted method to calculate BMDs, which has previously been used to assess nitrosamine potency (Chain et al. 2023; Johnson et al. 2021; Zeilmaker et al. 2010). PROAST allows benchmark doses to be calculated either via a graphical user interface, a command line interface, or via a web service. All of these methods however only allow BMDs to be calculated on an individual per-response basis and require significant manual input for each benchmark dose calculated. Due to the large number of benchmark doses necessary for this analysis (1175), manual calculation is impractical. This is especially true if benchmark dose modeling is to be used more widely on a mass scale. Instead, an automated approach is necessary.

To facilitate automated analysis, a wrapper was developed that sits on top of PROAST and provides a simple REST API allowing benchmark doses to be calculated for large datasets automatically.

The system is built around the command line interface to PROAST. To calculate a benchmark dose, a post request is made containing the individual dose–response data to be analysed. After this, the following steps are taken automatically (Fig. 3).

1. A copy of the posted data is saved to a temporary file in the correct format to be read by PROAST.
2. A bash script is generated containing the commands necessary to run the command line version of PROAST in R, as well as input the desired settings and save the results to a temporary output file (in JSON format).
3. The bash script is executed producing a copy of the analysis results.
4. The results are read from the output file and the temporary files are cleaned up.
5. The results are returned to the user.

The system is containerised as a docker image (Docker Inc. 2024) for ease of deployment and supports concurrent calculation of multiple benchmark doses to facilitate the analysis of large volumes of data. The API itself is written in Python 3.10 (Van Rossum and Drake Jr 2009) using the fastapi package. In line with standard practices, benchmark doses were calculated from quantal data using model averaging (where multiple models are fitted and the results are combined; logit and probit were excluded as recommended; no covariates were selected) using a 10% (extra risk) benchmark response size. Confidence intervals were set to 90% coverage and calculated from 200 bootstrap runs.

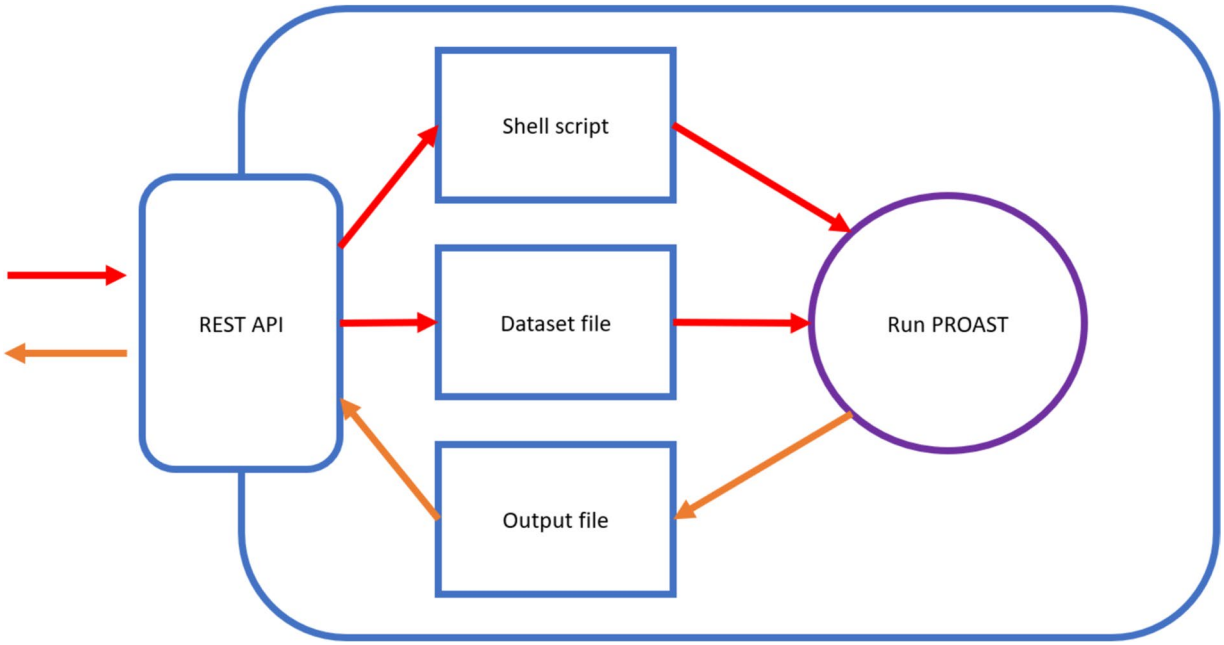

Fig. 3 Design and functionality of the PROAST wrapper. Predictions are made via a REST API. The API generates a script that runs PROAST with the appropriate settings. This then runs PROAST on the provided data and the results are passed back to the user. Red arrows: flow of input data. Orange arrows: flow of output data

Data analysis

Compound-level BMDs were calculated using the standard method whereby the harmonic mean of the potency estimates from the most sensitive species are used to generate the compound-level estimate (Thresher et al. 2019). To verify the method and dataset collection, compound-level TD_{50}s were also calculated based on TD_{50}s directly generated from the collected dataset. As methodology, an updated version of the one described in Thresher et al. (Thresher et al. 2019) was used (manuscript under preparation). These were able to replicate the existing summary TD_{50}s from LCDB. As TD_{50}s represent a maximum likelihood estimate, as well as a median predicted value under the TD_{50} model, whereas the commonly used BMDL supplied by PROAST has a lower confidence interval, it would not be fair to directly compare the BMDLs given by PROAST with their equivalent TD_{50}s. While PROAST does not provide a maximum likelihood estimate for the BMD, when model averaging is employed the confidence interval is calculated based on a sampling method and the samples are returned. We used the median value of the sampled BMDs as a point estimate for the benchmark dose. This is comparable to the TD_{50} in that both estimates represent a median potency estimate, it is also consistent with the confidence intervals supplied by PROAST which will converge on the median value as the confidence interval is reduced.

When experimental dosing durations do not equal the tested animals' lifespan, an adjustment is typically made to TD_{50}s to account for differences in dose duration (Peto et al. 1984)(see Eq. (4)). While there is no similar standard for lifespan adjustments for BMDs, the TD_{50} adjustment is model agnostic, depending only on dose duration and lifespan. Therefore, to provide an equivalent comparison, benchmark doses were adjusted using the standard TD_{50} lifespan adjustment.

$$TD_{50adjusted} = TD_{50} \times \left(\frac{\text{study duration}}{\text{lifetime}} \right)^2 \quad (4)$$

Results

Data preparation

While it is almost always possible to calculate a TD_{50} or benchmark dose in many cases the values are not meaningful, in that they do not represent a biologically relevant response. Typically for TD_{50}s this occurs when there is no response (no carcinogenic effects were observed in the respective experiments) and an arbitrary large number will be returned by the TD_{50} model. In the case of the TD_{50} model an AIC test (Akaike 1973) is used to identify cases where no response is present, and these values are then excluded from further analysis. Benchmark doses calculated using the current version of PROAST do not provide an explicit response curve necessary for calculating a log-likelihood, making an equivalent test impossible. Additionally, it is possible for PROAST to fail to converge on some responses. While a BMD analysis results in BMDs and BMDLs derived from the dose–response data, it also serves as a tool for evaluating the (statistical) quality of the data. Criteria for data rejection are currently under discussion in various groups, including EFSA (Benford et al. 2010). It has been suggested that the ratio of BMDL to BMD might serve as a measure for this and that the data should be rejected if this ratio exceeds a particular value (e.g. 10 or 100, (Barlow et al. 2006)). Based on this, we used a convergence threshold of 100 for the ratio of BMDL to BMDU (more stringent than the BMDL to BMD ratio), with benchmark doses only being accepted if the upper confidence interval was less than 100 times the lower confidence interval as well as the response having passed the AIC test using the TD_{50} model. We thus categorise responses into 3 types depending on these criteria:

1. Not significant: responses which show no significant response (under the TD_{50} model), whether or not the benchmark dose model converged.
2. Un-converged: responses which were significant, but PROAST was unable to find a good solution for the benchmark dose.
3. Converged: responses which were significant and PROAST was able to find a good benchmark dose.

A breakdown of the categories is given in Table 1.

Comparison of individual responses

Analysis of individual responses was performed on the converged results. For the calculation of an acceptable intake (AI), linear extrapolation is commonly used from some point of departure. In line with this, the TD_{50} was linearly

Table 1 Comparison of model fits for the TD_{50} and BMD models

	BMD_{10}	TD_{50}
Converged	296	541
Un-converged	245	0
Not significant	634	634
Valid compounds	41	53

The BMD model was unable to find good potency estimates for all responses leading to a loss of 12 compounds when compared to the TD_{50} model. In total TD_{50}s were calculated for 53 and (converged) BMDs were calculated for 41 compounds

extrapolated downwards to a TD_{10} to give a comparable estimate to the BMD_{10}. It should be noted that as linear extrapolation is non-conservative relative to the true TD_{10} this will lead to an over-estimation (under-estimation of the potency) of approximately 31% (see Supplementary Material 1 for calculations) for the extrapolated TD_{10} compared to the true value under the TD_{50} model. However, while not an accurate estimate of the TD_{10} the extrapolated value is in line with common use under ICH M7 and so represents the relevant value in a regulatory context.

Both individual TD_{50}s and BMDs appear similarly distributed ($p = 0.85$ based on a Kolmogorov–Smirnov test of converged results), showing good agreement between the models on aggregate.

Individual responses showed a high level of agreement between the TD_{10} values and the BMD_{10}s, with an r^2 value of 0.94. A best-fit line was calculated using orthogonal least-squares on the logged values which found a close match with $\ln(BMD_{10}) = 0.96 \ln(TD_{10})—0.18$ (Fig. 4).

While the overall agreement was high there was significant variation around the best-fit line, with the log-difference in estimates appearing non-normal ($p = 0.002$) and heavy tailed with a mean error of 0.40 $\log_{10}$ units or approximately 2.5 times relative error (Fig. 5).

Potency comparisons summarised at compound-level

Summary values for each compound were calculated in line with the Carcinogenic Potency Database (CPDB) and Lhasa Carcinogenicity Database (LCDB) methods. Studies were grouped according to the species of test subject in which the compound was tested. For each study that showed an increase in tumour formation, the most potent TD_{50} was selected. The harmonic mean of the most potent TD_{50}s across each species was then used as a representative summary value for that compound within the species.

Compound summary TD_{10}s were able to be calculated for 53 of the 55 analysed compounds, with two compounds showing no significant responses. Summary BMDs were calculated for 41 compounds with the remaining compounds having no converged findings (Table 1).

While there was a high correlation between compound summary BMD_{10}s and TD_{10}s ($r^2 = 0.905$), the benchmark doses do not follow the same distribution as the TD_{10}s. A best-fit line using orthogonal least squares shows BMDs on average predict a lower potency but with higher variation in results [$\ln(BMD) = 1.35 \ln(TD_{10}) + 1.54$] (Figure 6).

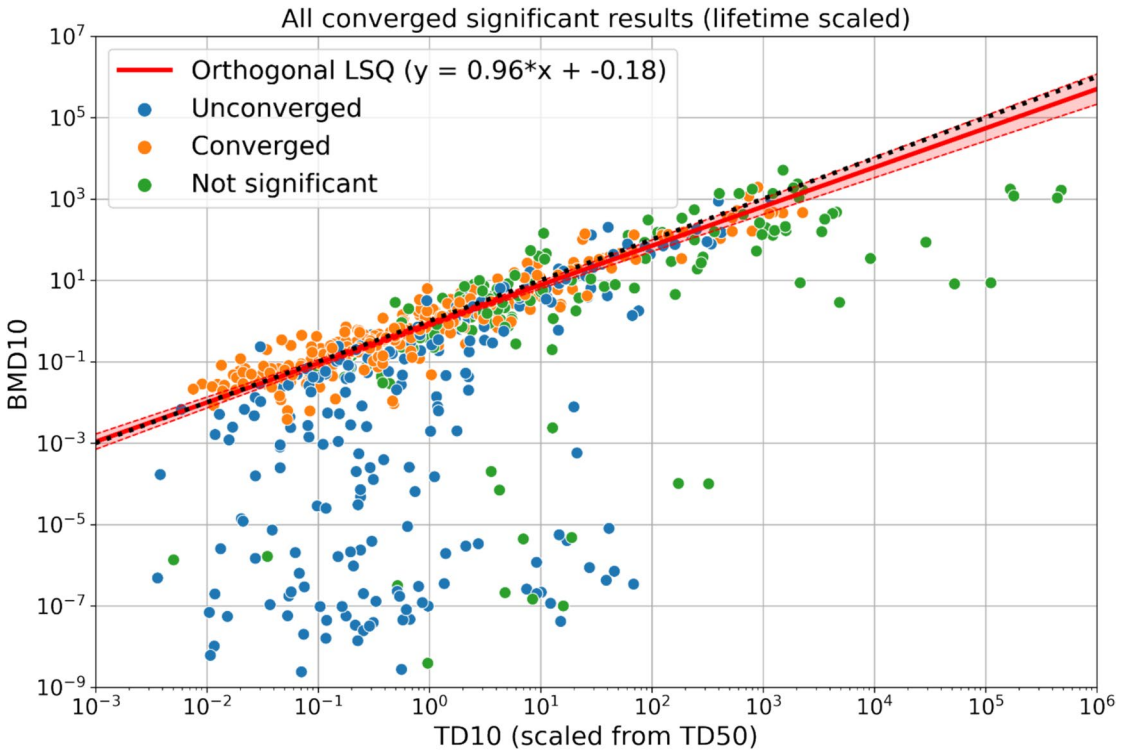

Fig. 4 Comparison of response-level potencies using BMD_{10} and TD_{10} values. The best-fit line is fitted against the (natural) log of these values for the converged responses. The importance of model convergence in robust potency estimation is illustrated by the poor correlation obtained from unconverged results relative to the agreement of the data to the best-fit line for the converged values. Potencies have been adjusted for experiment duration vs. lifetime

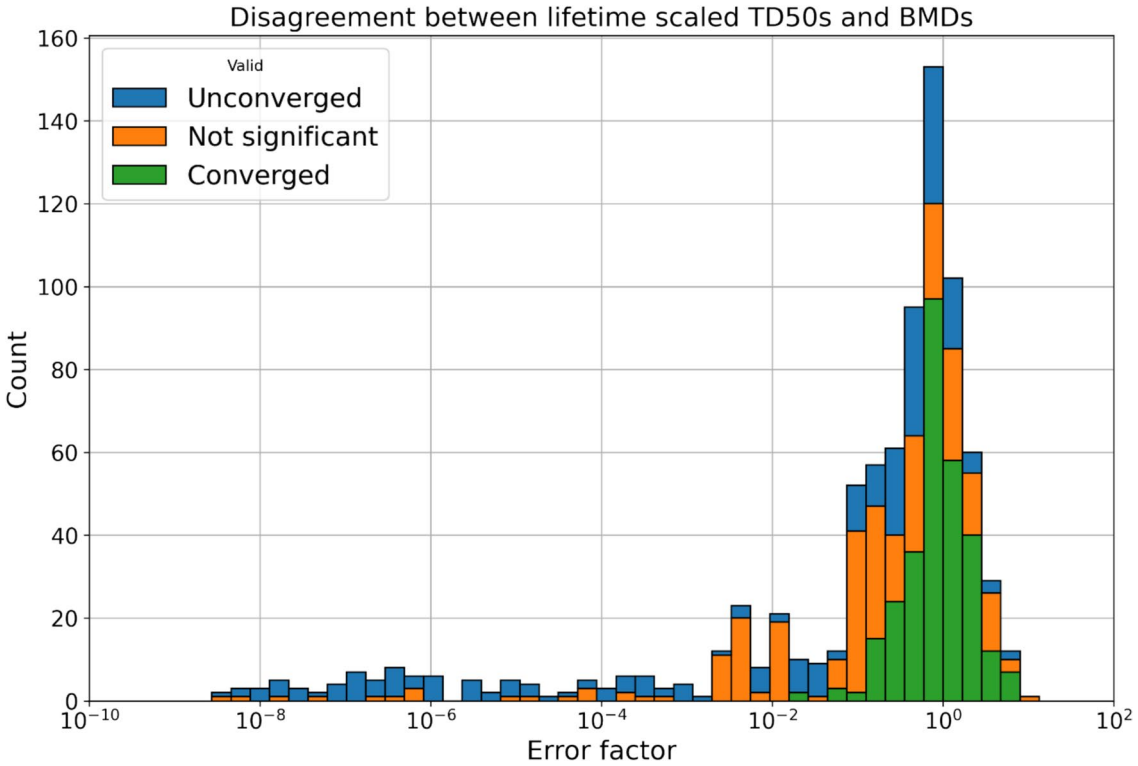

Fig. 5 Comparison of the difference between BMD_{10} and TD_{10} potency estimates. The distribution of the difference in potencies is heavy tailed with a mean error of 0.40 $\log_{10}$ units (approximately 2.5 times relative error). Additionally, the unconverged BMDs (BMD ratio > 100) are highly skewed towards low (high potency) estimates. "Error factor" is the fold-change of error between BMD_{10} and TD_{10} values, so a value of 10^2 would indicate a BMD 100 times greater than the corresponding TD10

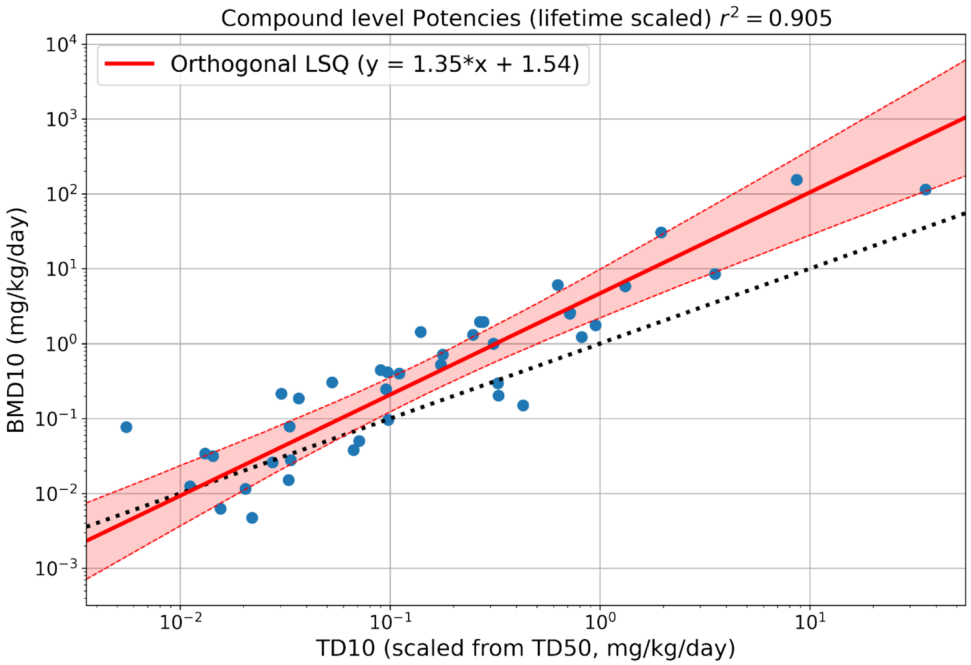

Fig. 6 Comparison of compound level potencies using the BMD_{10} and the TD_{10}. The best-fit line is fitted against the (natural) log potencies. Compound-level BMD estimates only account for the converged responses

This is supported by the overall distributions which both appear log-normal ($p = 0.476$ and $p = 0.471$ for log TD_{10}s and BMDs respectively). The TD_{10}s show a lower (more potent) mean value with $\log(TD_{10}) = -1.92$ compared to −1.02 for the $\log(BMD_{10})$, and a lower dispersion with a standard deviation of 1.88 compared to 2.46 for the BMDs (Fig. 7).

Acceptable intakes

One notable application of the benchmark dose is in the setting of acceptable intakes for nitrosamine impurities. Although partially superseded by the "Carcinogenic Potency Categorisation Approach" (CPCA) (EMA 2023; Kruhlak et al. 2024), a class limit of 18 ng has been proposed for nitrosamine intake based on the 5th percentile of nitrosamine TD_{50}s from LCDB. For comparison, we present an equivalent limit calculated on both the TD_{10}s and benchmark doses included in this analysis. Acceptable intakes have been calculated based on Eq. (5):

$$\mathrm{AI} = \frac{p5\{\mathrm{TD10}\}}{10000} \times 50\mathrm{kg} \text{ or } \mathrm{AI} = \frac{p5\{\mathrm{BMD10}\}}{10000} \times 50\mathrm{kg} \quad (5)$$

It is important to note that we derive a slightly different limit from the TD_{50}s to those previously published (Fine et al. 2023; Thomas et al. 2021). This likely occurred due to the exclusion of compounds for which we were unable to calculate a benchmark dose. Interestingly however the limit derived from the benchmark dose values is similar to that derived using the TD_{10}s (Table 2).

Although the benchmark doses were less potent estimates on average, the increased variability means that the lower extreme of the BMD_{10} distribution lies below that of the TD_{10} distribution. By chance, the cross-over point where both models produce the same potency estimate is at the 5.78th percentile, very close to the 5th percentile used to calculate the acceptable intakes. This suggests that while a class-level acceptable intake is unlikely to change significantly with the wider acceptance of benchmark doses, at least in the case of nitrosamines, individual estimates used for read-across may change significantly.

Table 2 Comparison of acceptable intakes based on the 5th percentile of the TD_{10} and BMD_{10} distributions

	5th percentile (mg/kg/day)	Acceptable intake (ng/day)
TD_{10}	0.0065	33
BMD_{10}	0.0062	31

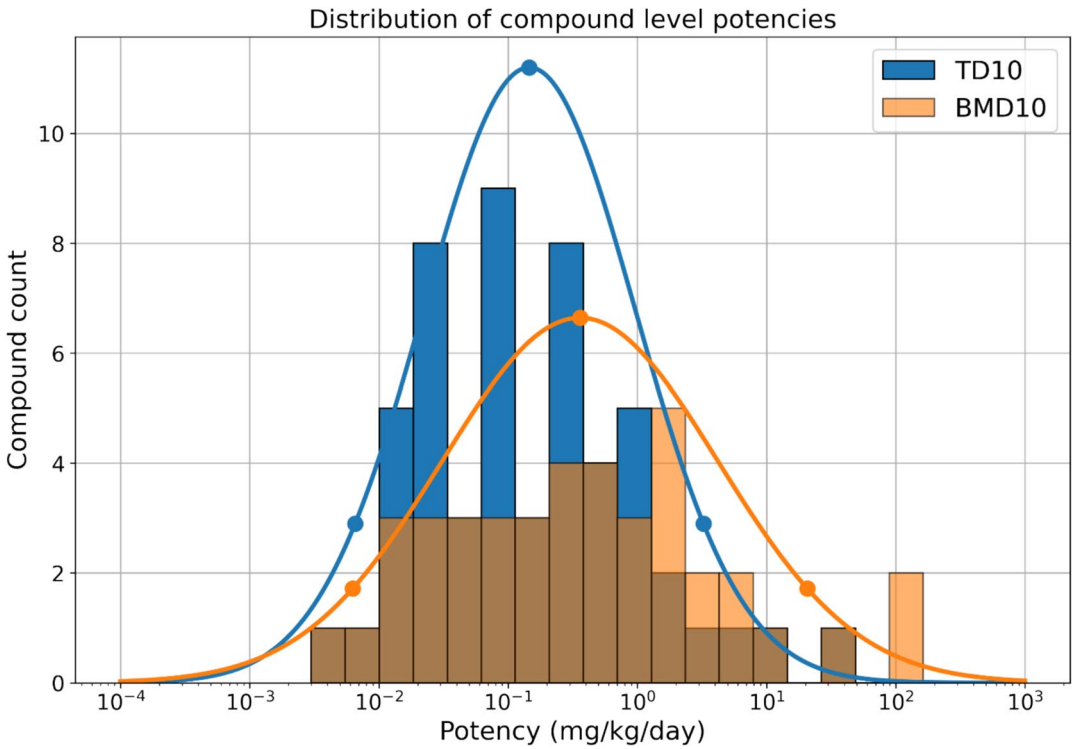

Fig. 7 The distributions of compound-level potencies using TD_{10} or BMD_{10} methods. The benchmark dose approach yielded, on average, lower potency estimates (i.e., higher values) coupled with greater variation. Points show the 5th, 50th, and 95th percentiles

Discussion

While undoubtedly successful, the TD_{50} model takes a very simplified view of the expected response and was never intended to be used on a large scale and without consideration of the underlying response, as made clear by the following quote from Peto et al. in the original paper (Peto et al. 1984) proposing the TD_{50} model (emphasis added).

Different statisticians would undoubtedly devise different statistical methods for estimating the TD_{50}. One simple way would be to calculate for each group the probability P of remaining tumorless and to plot a graph of these probabilities against dose rate. With P plotted on a log_2 scale, there will then be a unit change in log_2 P as we go from zero dose to the TD_{50}. It may be that such a graph will yield approximately a straight line, for this is predicted by certain rather simple multistage models for cancer induction. However, other equally plausible multistage models do not predict straight lines, and so the expectation that the line might be straight must not distort the interpretation of the actual plotted data.

Despite this, the TD_{50} model has been broadly accepted as the standard method for calculating carcinogenic potency and has been applied on mass in the widely used CPDB and later the LCDB as well as embedded in regulations such as ICH M7 as the default approach. The widespread adoption of the TD_{50} model may in part be due to the simplicity of calculating a TD_{50} compared to other more generalisable methods such as the benchmark dose modelling. The ability to apply more sophisticated models on a larger scale—as demonstrated here—is an important step towards the adoption of more robust methods.

Benchmark dose modelling has been shown to provide more relevant potency values than the TD_{50} approach due to the use of a smaller effect-size (defaulting to 10% rather than 50%) as well as the increased flexibility of the models enabling different response types to be fitted effectively. Furthermore, single metrics such as the TD_{50} don't have a measure of precision, and confidence intervals are needed. For these reasons, BMD has been recommended by EFSA, EMA, RIVM and other regulatory/public health organisations for use in certain situations (Chain et al. 2023; Slob et al. 2014).

The comparisons presented in this work show that while usage is unlikely to affect safe nitrosamine class limits there is considerably more variation in potency than accounted for by the TD_{50} model, with the majority of compounds showing lower potency than previously expected.

Read-across

Using the most relevant data we can calculate the BMD with the most precise potency estimate for inter-molecule comparison or categorisation. Table 3 shows the values for several key nitrosamine compounds that have been used for read-across assessments, i.e., from studies generally considered to be very robust. For these compounds, very little change in potency is observed: all are within a threefold range (less than half an order of magnitude)—and in the context of the inherent variability in a biological response, in an overall landscape wherein potencies can change by four orders of magnitude (Thresher et al. 2020), this is an acceptable variation: N-nitrosodimethylamine (NDMA, N-nitrosotetrahydropyridine (NTHP), N-nitrosodiethanolamine (NDELA) and N-nitrosodiphenylamine (NDPhA) are all less potent by benchmark dose than by TD_{10}, and N-nitrosomorpholine (NMOR), 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), N-nitrosopyrrolidine (NPYR), N-nitrosodiethylamine (NDEA) and N-nitrosopiperidine (NPIP) are considered to be of marginally higher potency by benchmark dose than by TD_{10}. It is important also to note that this set of compounds covers the four orders of magnitude mentioned above, ranging from NDEA, one of the most potent, to the

Table 3 Nitrosamines relevant for read-across with associated potency changes

Compound	CAS No	TD_{10} (mg/kg/day)	Lifetime BMD_{10} (mg/kg/day)	Difference (mg/kg/day)	fold change
NDELA	1116-54-7	0.72	2.576	1.856	2.58
NDPhA	86-30-6	35.7	113.7	78.153	2.20
NDMA	62–75-9	0.033	0.079	0.046	1.37
NTHP	55556-92-8	0.014	0.032	0.017	1.21
NMOR	59–89-2	0.027	0.026	–0.002	0.963
NNK	64091-91-4	0.071	0.05	0.021	0.704
NPYR	930-55-2	0.328	0.202	–0.125	0.616
NDEA	55-18-5	0.016	0.006	–0.009	0.375
NPIP	100-75-4	0.43	0.15	–0.279	0.349

Note that compound level TD_{10}s were calculated based on the data for which BMDs could be calculated as well, for the purpose of fair comparison to the respective BMD_{10}s in this table. They may differ from published values and should not be used as points of departure (POD) for acceptable intake extrapolation

least potent NDPhA—and there is no change in the trend across the range of potencies.

BMD comparisons across endpoints

As BMD modelling is applicable to wide-ranging toxicological endpoints and not just carcinogenicity studies, it is well-suited for the provision of quantitative potency estimations supporting quantitative adverse outcome pathways (AOPs) and multi-compound relative potency comparisons. There is also a growing body of work utilising cross-endpoint BMD correlations to establish the relationships between shorter-term in vivo genetic toxicity tests and the outcomes of the rodent cancer bioassay (Chepelev et al. 2023; Soeteman-Hernández et al. 2016). Quantitative potency determinations achieved used BMD modelling can also support genotoxic mechanism-of-action assessments and harness BMD confidence interval information from in vivo genetic toxicity tests to calculate acceptable intakes in place of cancer bioassay data (Johnson et al. 2021; Wills et al. 2015). In this way, implementing a BMD-derived, relative potency-based approach for nitrosamine assessment could enable the setting of acceptable intake limits from in vivo mutation dose–response data. This would expand beyond the current CPCA strategy and could allow for the adjustment of CPCA categories based on structure–activity relationships alone. It might also enable the derivation of higher acceptable intake limits beyond current CPCA limits for nitrosamines such as large nitrosamine drug substance-related impurities (NDSRIs) where quantitative potency determinations from genetic toxicity dose–response information demonstrate low or no detectable mutagenic potency.

Conclusion

Benchmark dose modelling offers a more flexible family of models capable of describing the sigmoidal nature of many dose–response relationships with a higher degree of accuracy than the simple TD_{50} approach. Using a large nitrosamine dataset as a case study, we compared the impact of moving to exposure limits set from BMD modelling instead of the TD_{50}. Whereas differences in individual potency estimates were noted between the two approaches for some compounds, key metrics such as the 5th percentile used for calculating acceptable intakes were not greatly affected. Moreover, potency estimations for nitrosamine compounds relevant to read-across did not vary by more than a factor of 3. Overall, this suggests that a move to BMD-derived limits is achievable without significant disruption to existing limits whilst placing future reliance on a far more robust methodology.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00204-024-03951-8.

Acknowledgements The authors want to thank Leonardo Allain (MSD) for his support and valuable comments on the manuscript.

Author contributions DJP: data curation, writing—original draft, writing—review & editing. RT: lead data analyst, data curation, writing—original draft, writing—review & editing. AT: data curation, writing—review & editing. GJ: organisation of conduct of the study, critiqued the output for important intellectual content. JS: writing—original draft, writing—review & editing. JW: conceptualisation, writing—review & editing. All authors reviewed the results and approved the final version of the manuscript.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest The authors declare no conflict of interest, financial or otherwise. GJ is a consultant who evaluates the risks posed by pharmaceutical impurities. His clients did not influence the content of this manuscript.

Ethical approval The manuscript does not contain clinical studies or patient data.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. Springer, New York

Barlow S, Renwick AG, Kleiner J et al (2006) Risk assessment of substances that are both genotoxic and carcinogenic report of an International Conference organized by EFSA and WHO with support of ILSI Europe. Food Chem Toxicol 44(10):1636–1650. https://doi.org/10.1016/j.fct.2006.06.020

Benford D, Bolger PM, Carthew P et al (2010) Application of the margin of exposure (MOE) approach to substances in food that are genotoxic and carcinogenic. Food Chem Toxicol 48(Suppl 1):S2-24. https://doi.org/10.1016/j.fct.2009.11.003

Burns MJ, Ponting DJ, Foster RS et al (2023) Revisiting the landscape of potential small and drug substance related nitrosamines in pharmaceuticals. J Pharm Sci 112(12):3005–3011. https://doi.org/10.1016/j.xphs.2023.10.001

Chain EPoCitF, Schrenk D, Bignami M, et al. (2023) Risk assessment of N-nitrosamines in food. EFSA J 21(3):e07884 https://doi.org/10.2903/j.efsa.2023.7884

Chepelev N, Long AS, Beal M et al (2023) Establishing a quantitative framework for regulatory interpretation of genetic toxicity dose-response data: margin of exposure case study of 48 compounds with both in vivo mutagenicity and carcinogenicity dose-response data. Environ Mol Mutagen 64(1):4–15. https://doi.org/10.1002/em.22517

Docker Inc. (2024) Docker manuals. https://docs.docker.com/manuals/

Edler L, Poirier K, Dourson M et al (2002) Mathematical modelling and quantitative methods. Food Chem Toxicol 40(2–3):283–326. https://doi.org/10.1016/s0278-6915(01)00116-8

EMA (2023) Appendix 2 to Questions and answers for marketing authorisation holders/applicants on the CHMP Opinion for the Article 5(3) of Regulation (EC) No 26/2004 referral on nitrosamine impurities in human medicinal products - Carcinogenic Potency Categorisation Approach for N-nitrosamines. In. https://www.ema.europa.eu/en/documents/other/appendix-2-carcinogenic-potency-categorisation-approach-n-nitrosamines_en.pdf Accessed OCT 6, 2023

European Food Safety Authority (2005) Opinion of the Scientific Committee on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic. EFSA J 3(10):282. https://doi.org/10.2903/j.efsa.2005.282

Fine J, Allain L, Schlingemann J, Ponting DJ, Thomas R, Johnson GE (2023) Nitrosamine acceptable intakes should consider variation in molecular weight: the implication of stoichiometric DNA damage. Regul Toxicol Pharmacol 145:105505. https://doi.org/10.1016/j.yrtph.2023.105505

ICH (2017) ICH guideline M7(R1) on assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk.

Johnson GE, Dobo K, Gollapudi B et al (2021) Permitted daily exposure limits for noteworthy N-nitrosamines. Environ Mol Mutagen 62(5):293–305. https://doi.org/10.1002/em.22446

Kroes R, Galli C, Munro I et al (2000) Threshold of toxicological concern for chemical substances present in the diet: a practical tool for assessing the need for toxicity testing. Food Chem Toxicol 38(2–3):255–312. https://doi.org/10.1016/s0278-6915(99)00120-9

Kroes R, Renwick AG, 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(1):65–83. https://doi.org/10.1016/j.fct.2003.08.006

Kroes R, Kleiner J, Renwick A (2005) The threshold of toxicological concern concept in risk assessment. Toxicol Sci 86(2):226–230. https://doi.org/10.1093/toxsci/kfi169

Kruhlak NL, Schmidt M, Froetschl R et al (2024) Determining recommended acceptable intake limits for N-nitrosamine impurities in pharmaceuticals: development and application of the Carcinogenic Potency Categorization Approach (CPCA). Regul Toxicol Pharmacol 150:105640. https://doi.org/10.1016/j.yrtph.2024.105640

Lhasa Limited (2023) Lhasa carcinogenicity database. https://www.lhasalimited.org/products/lhasa-carcinogenicity-database.htm

Müller L, Mauthe RJ, Riley CM et al (2006) A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity. Regul Toxicol Pharmacol 44(3):198–211. https://doi.org/10.1016/j.yrtph.2005.12.001

Munro IC (1990) Safety assessment procedures for indirect food additives: an overview. Report of a workshop. Regul Toxicol Pharmacol 12(1):2–12. https://doi.org/10.1016/s0273-2300(05)80042-x

Peto R, Pike MC, Bernstein L, Gold LS, Ames BN (1984) The TD50: a proposed general convention for the numerical description of the carcinogenic potency of chemicals in chronic-exposure animal experiments. Environ Health Perspect 58:1–8. https://doi.org/10.1289/ehp.84581

Slob W, Bakker M, Biesebeek J, Bokkers B (2014) Exploring the uncertainties in cancer risk assessment using the integrated probabilistic risk assessment (ipra) approach. Risk Anal. https://doi.org/10.1111/risa.12194

Soeteman-Hernández LG, Johnson GE, Slob W (2016) Estimating the carcinogenic potency of chemicals from the in vivo micronucleus test. Mutagenesis 31(3):347–358. https://doi.org/10.1093/mutage/gev043

Thomas R, Thresher A, Ponting DJ (2021) Utilisation of parametric methods to improve percentile-based estimates for the carcinogenic potency of nitrosamines. Regul Toxicol Pharmacol 121:104875. https://doi.org/10.1016/j.yrtph.2021.104875

Thresher A, Gosling JP, Williams R (2019) Generation of TD(50) values for carcinogenicity study data. Toxicol Res 8(5):696–703. https://doi.org/10.1039/c9tx00118b

Thresher A, Foster R, Ponting DJ, Stalford SA, Tennant RE, Thomas R (2020) Are all nitrosamines concerning? A review of mutagenicity and carcinogenicity data. Regul Toxicol Pharmacol 116:104749. https://doi.org/10.1016/j.yrtph.2020.104749

Van Rossum G, Drake Jr FL (2009) Python 3 reference manual. CreateSpace. https://docs.python.org/3/reference/index.html

Wills JW, Johnson GE, Doak SH, Soeteman-Hernández LG, Slob W, White PA (2015) Empirical analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings and support MOA determinations. Mutagenesis 31(3):255–263. https://doi.org/10.1093/mutage/gev085

Zeilmaker MJ, Bakker MI, Schothorst R, Slob W (2010) Risk Assessment of N-nitrosodimethylamine formed endogenously after fish-with-vegetable meals. Toxicol Sci 116(1):323–335. https://doi.org/10.1093/toxsci/kfq093