ARTICLE

Generation of TD$_{50}$ Values for Carcinogenicity Study Data

Andrew Thresher, *[a] John Paul Gosling [b] and Richard Williams [a]

Carcinogenic potency is a key factor in the understanding of chemical risk assessment. Measures of carcinogenic potency, for example TD$_{50}$, are instrumental in the determination of metrics such as the threshold of toxicological concern (TTC), acceptable intake (AI) and permitted daily exposure (PDE), which in turn impact on human exposure. The Carcinogenic Potency Data Base (CPDB) has provided a source of study information, complete with calculated TD$_{50}$ values. However, this is no longer actively updated. An understanding of carcinogenic potency, which can be derived from dose-response data, can be used as part of human risk assessments to generate safety thresholds under which cancer risk is judged to be minimal. The aim of this paper is to produce a transparent methodology for calculating TD$_{50}$ values from experimental data in a manner consistent with the CPDB. This was then applied across the same data as used in the CPDB and analysis done on the correlation with the CPDB TD$_{50}$ values. While the two sets of values showed a high level of correlation overall, there were some significant discrepancies. These were predominantly due to a lack of clarity in the CPDB methodology and inappropriate use of a linear model in TD$_{50}$ calculation where the data was not suitable for such an approach.

1. Introduction

There is a wealth of data on chemical carcinogenicity from animal studies available in the published literature. Until 2007, this was actively collated and curated as part of the Carcinogenic Potency Data Base (CPDB) project under the guidance of Lois Gold (Carcinogenic Potency Project).[1] Sources included long-term carcinogenicity data from both general literature and through collaboration with the US National Toxicology Program (NTP) (Chemical Effects in Biological Systems).[2]

These data are routinely used for chemical safety assessments, where carcinogenicity is (understandably) one of the endpoints considered. This requires an understanding of carcinogenic potency to calculate exposures at which carcinogenic risk can be assumed to be acceptable, which is likely to differ based on the application of the chemical being considered, as part of a risk-benefit comparison or due to differences in exposure routes, for example.

Historically, the metric used to determine carcinogenic potency has been the TD$_{50}$, defined as the dose required to halve the probability of a subject remaining without tumours throughout a lifetime of exposure. TD$_{50}$ values were included in the CPDB and subsequently used to derive the threshold for toxicological concern (TTC) for carcinogens,[3] which is widely used as a pragmatic safety threshold for chemicals lacking adequate experimental data. However, alternative approaches to TTC calculation and the use of alternative metrics for carcinogenic potency have been recommended more recently.[4]

Despite the possibility of using alternative metrics, the application of TD$_{50}$ values for risk assessments in international regulations and industrial practice remains widespread. For example, in the ICH M7 guidance for genotoxic impurities in pharmaceuticals, linear extrapolation from TD$_{50}$ values is cited as the default method for calculating a compound-specific acceptable intake where adequate (positive) carcinogenicity data is available. Moreover, thresholds for human exposure to 14 genotoxic impurities were calculated, 10 of which were derived from the linear extrapolation of TD$_{50}$ values that were either present in CPDB or "calculated from published studies using the same method as in the CPDB".[5] A similar process utilising TD$_{50}$ values was used to derive acceptable intakes for the (only) two mutagenic impurities described in a recent pharmaceutical industry study, as well as a class-specific limit for alkyl bromides.[6]

a. Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, West Yorkshire, LS11 5PS, UK
b. School of Mathematics, University of Leeds, Leeds, UK

The calculation of TD_{50} values is described in some detail in several publications, the most well-known being Cox (1972)[7], Peto *et al.* (1984)[8] and Sawyer *et al.* (1984)[9]; however, these methods require the use of 'lifetable' data where tumour incidence is tracked over time. Such data are rarely available given current experimental protocols. In the more common case where only tumour incidence data at the terminal sacrifice is available, the CPDB used a 'summary' methodology that is far less detailed in the description of incidences over time.[10]

The aim of the current work was to reproduce the methodology by which TD_{50} values could be generated from carcinogenicity study data using modern statistical tools and knowledge, but in a manner consistent with the approach used by Gold *et al.* (1984).[11] Once complete, it would then be possible to make the resultant scripts publicly available. This has two benefits. Firstly, it will provide a documented and reproducible method to generate TD_{50} values which can be used for risk assessments for genotoxic carcinogens, such as those already described. Secondly, the creation and sharing of scripts allows the automated generation of TD_{50} values to accelerate benchmarking against new metrics to describe points of departure, such as benchmark dose.

The first step in the work involved the digestion of the published studies on TD_{50} calculation, many of which were published over 20 years ago, and the implementation of scripts in the statistical package R to reproduce the calculations.[12] Then, an extensive validation was conducted to compare the TD_{50} values generated using our R-script to those reported in CPDB.

2. Materials & Methods

2.1. Data

To allow for a practical comparison with the CPDB TD_{50} values, long-term carcinogenicity study data was obtained from the CPDB website. Making use of the same data as was used in the generation of the CPDB minimised the potential for complicating factors which may arise when comparing the newly generated TD_{50} values with the CPDB TD_{50} values.

Chemical information (that is, CAS registry number[13] and common name) for each test substance was cross-referenced with the Vitic_Lhasa 2016.1 database (Lhasa Limited. Vitic Nexus)[14]. Any instances where the same chemical entity appeared multiple times as separate test

substances in the CPDB were merged with relevant information recorded with the associated studies rather than the chemical compound. For example, two batches of a single chemical compound may have been tested separately, with each batch displaying a different chemical purity. Rather than recording these as unrelated entities (as was done in CPDB), these have been assigned the same test substance ID with the relevant purity information being associated with the individual studies. In doing this the total number of unique test substances for which carcinogenicity data was available was reduced from 1,547 to 1,529.

The data set contained an 'opinion' column, which the CPDB documented as being an indication of the original study authors' tumourigenic activity call. Data were removed for any tumour type-tumour site combination which was considered to be negative, as these data are not amenable to TD_{50} modelling due to a lack of tumourigenic affect.

The CPDB also assigned each dose-response set with a 'curve' notation (*, \, /, | or 0), indicating the shape of the dose-response curve with regard to a linear correlation. Table 1 details the nature of each curve notation together with its prevalence within the data set as a whole. Those assigned a '0' curve either display no dose-response relationship or consist of a single treatment group. As such all data assigned a '0' curve were removed from the data set.

2.2. Statistical Models

Two statistical models were used to generate the TD_{50} estimates and associated confidence intervals. The choice of model for each data set depended on whether the full lifetable data were available or if the end-of-experiment statistics were recorded. Both models were coded in the R statistical language with optimisations for the calculations of the maximum likelihoods and confidence intervals using the Broyden–Fletcher–Goldfarb–Shanno algorithm[15] implemented in the built-in *optim* function.

The first model was used when full lifetable data were available, and it followed the estimation process as laid out in Sawyer *et al.* (1984).[9] Here, a form of proportional hazards model was utilised to model the effect of changing dose on the probabilities of not forming a tumour over time. In this modelling approach, account is taken of when the tumours occurred and the number of animals at risk over each time period in any given

experiment. The estimates for the TD_{50} are based upon maximum likelihood estimates stemming from a generalised linear model with a binomial likelihood. This form is used because a linear model can be used to capture the effect of changing dose on the probability of getting a tumour for each observation period.

The second model was used when only the final numbers of animals exposed and which developed tumours were reported for each dose level. In this case, the data are not as informative because it is not known when the tumours occurred in the experiment. This model is again based on a binomial likelihood, but the time element is removed. Again, a linear model is used to capture the effect of dose on the probability of tumours occurring. A description of this method is given in the appendix of Peto *et al.* (1984).[8]

For both models, the lead of Peto *et al.* (1984)[8] was followed by applying an extrapolation factor to the TD_{50} estimates to accommodate the fact that the experiments were run at a different timespan to the timespan of interest: that is, the lifespan of the animal. The particular multiplier for this purpose was the square of the ratio of the length of the experiment to the standard lifespan for mice, which was estimated at 104 weeks for rodent species, 234 weeks for tree shrews, 520 weeks for bush babies, 572 weeks for dogs and 1040 weeks for monkey species.

When fitting each model, a test was performed to determine if a constant relationship between the probability of tumours occurring and dose better accorded with the data than the linear dose relationship that is integral to both models. In practice, a likelihood ratio test was conducted to establish if the model with a positive linear relationship was better than a constant fit. If a constant relationship performs better, then it is probable that the linear model is not suitable to capture the experimental results.

2.3. Analysis

Analysis was conducted in the KNIME Analytics Platform (version 3.1.2.).[16] The TD_{50} models were implemented as part of a Knime workflow. The script was encapsulated into an R snippet node which was positioned within a loop set to iterate across the dose-response data for each tumour-tissue pairing. In this manner it was possible to run the TD_{50} calculation script across a large number of dose-response curves.

To reproduce the workings of Gold *et al*. (1986)[10], the data set was split according to the curve notation that had been assigned in the CPDB. This separated the data into a number of groups and each of these groups was processed as recommended by the CPDB documentation, with the exception of the '0' curve group, which was removed (Table 1).

Table 1: Details of curve notation used to classify dose-response data in CPDB

Group	Definition	Processing required	Number of tumour-tissue data sets [a]
*	Two or more dose groups showing a linear curve	None	8,180
\	Two or more dose groups showing a downward departure from linearity	Only the dose level displaying the highest tumour incidence was retained for TD_{50} calculation	543
/	Two or more dose groups showing an upward departure from linearity	None	672
Z	Two or more dose groups showing upward or downward departure from linearity	A column in the CPDB data file indicated the number of dose groups used when generating the TD_{50}. The highest tested dose groups were removed until this number of dose groups was reached	1,147
0	No dose effect or only a single dose group	Removed from the data set	12,179

[a] Number within the whole data set before the removal of negative authors' call rows

ARTICLE

The CPDB also performed a summation step where a single TD_{50} value was generated per compound per species. We reproduced this per compound per species TD_{50} value using the same method as CPDB. Data from all tissue/tumour data sets, excluding those which did not demonstrate a positive tumourigenic response, which were assigned a '0' curve notation, or which generated outlying TD_{50} values, were gathered and the most potent TD_{50} value within each study selected. These values were then grouped according to test animal species and the harmonic mean calculated, giving a single TD_{50} value per species per compound. In order to also produce a single, overall TD_{50} value per compound this method was modified by removing the species grouping and so calculating the harmonic mean from all applicable studies for each compound.

The correlation between the R script-generated TD_{50} values and those published within the CPDB was assessed using the Pearson correlation coefficient. This was calculated using the .corr() method in the Python programming language.[17]

3. Results

3.1. Distribution of TD_{50} values

There is some doubt regarding the nature of the data used by Gold *et al.* (1984)[11] in relation to the availability of lifetable data (see section 4). Subsequently, all data sets were processed using the second method described in section 2.2 (that is, the method reproducing the 'summary' methodology, see supplementary appendix A). Henceforth, all reference to R script-derived TD_{50} values refer to those generated using this method.

After removal of the dose-response groups with either a negative authors' call or a '0' curve notation, a total of 8,969 data points remained, each representing a unique combination of study ID, tumour site and tumour type. TD_{50} values were successfully calculated using the R script described above for 8,956 of these (leaving 13 data points with no R-script generated TD_{50} value). The distributions of both the published CPDB and R script-generated TD_{50} values are shown in Figure 1.
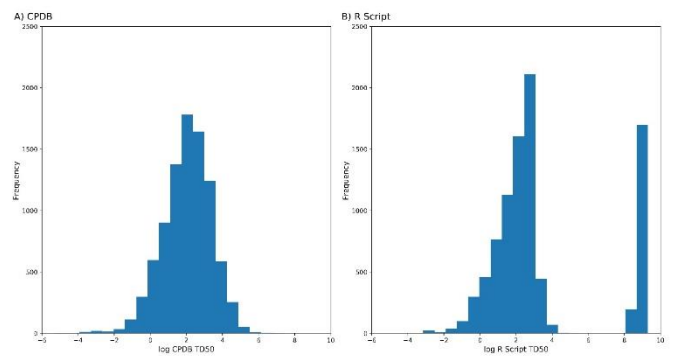

Figure 1: Distribution of log TD_{50} values a) downloaded from the CPDB website and b) calculated by the R script

The calculated TD_{50} values show a number of unusually high values which are isolated from the main distribution curve (log R script TD_{50} greater than 6). A similar set of outliers appears in the log CPDB TD_{50} distribution but only in those data points removed due to a negative authors' call or '0' curve notation. These values represent impractically high dosing levels when applied. For example, a log TD_{50} value of 6 would equal a dose of 1,000,000 mg/kg, or a dose equivalent to the mass of the test subject. It appears that these values are generated where the methodology is unable to calculate a valid TD_{50} value from the available dose-response data. This generally occurred when the incidence of tumours remained low over all dose groups (for example, 0/28, 0/29, 0/29, 1/27, 1/27, 2/29 and 2/29 animals with tumours/total animals) or where the number of animals in each dose groups was low (for example, 0/7, 0/9, 0/8, 0/9, 2/5 animals with tumours/total animals). Both these examples were classified as 'positive' by the study author. In the latter case, it has been suggested that data such as this, where the dose response groups are too small to generate a point of departure metric, should be excluded from such calculations.[4] The former case is more difficult to assess as part of a large validation exercise because it is possible that unusual tumours may result in a positive classification despite the low incidence; that is, there may be biological relevance that usurps either a human or statistical evaluation of the dose response data alone.

Removal of data sets where TD_{50} was calculated to be greater than 6, henceforth referred to as 'outliers', leaves 7,061 data sets with valid R script TD_{50} values. While the distribution of these is broadly similar to that of the CPDB values, it is also clear that the R script is not generating TD_{50} values which are an exact match of those from the CPDB. The median log TD_{50} for the R script-generated values (after removal of the outliers) was 2.11 compared to 2.13 for the CPDB values, with standard deviations of 1.09 and 1.35, respectively.

This can also be observed when comparing the overall, per compound TD_{50} values obtained by analysing the data for all studies related to each compound as described in section 2.3. The distribution of these overall TD_{50} values is shown in Figure 2.
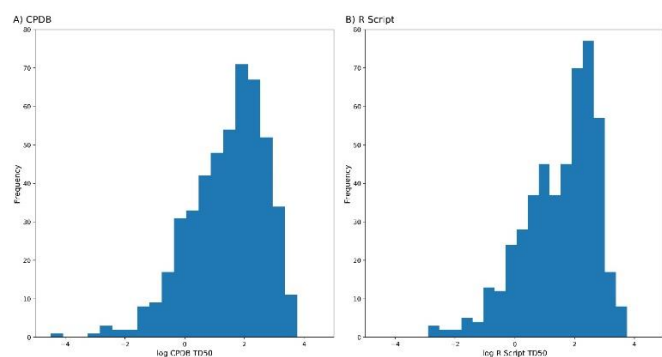

Figure 2: Distribution of per compound log TD_{50} values a) calculated from CPDB TD_{50} values and b) calculated from the R script TD_{50} values

3.2. Correlation of TD_{50} values

After removing the data sets which were 1) considered negative, 2) displayed '0' curve notation, 3) had no R script-generated TD_{50} value, or 4) where the R script generated an outlier TD_{50} value, a total of 7,061 TD_{50} values remained, covering 664 compounds. This still includes compounds with no author call, or an equivocal call (removal of these resulted in per compound calls for 486 test substances). The correlation between the CPDB and R script-generated TD_{50} values is illustrated in Figure 3. The Pearson coefficient generated values of 0.926 when comparing the individual dose-response TD_{50} values (that is, those related to specific site and tumour combinations, Figure 3a) and 0.979 when comparing the overall, per compound TD_{50} values (Figure 3b).

These indicate that although the calculated values do not exactly match those published by the CPDB, a good correlation between the two sets of values remains.

Potential factors which may have led to these differences are discussed in section 3.4.
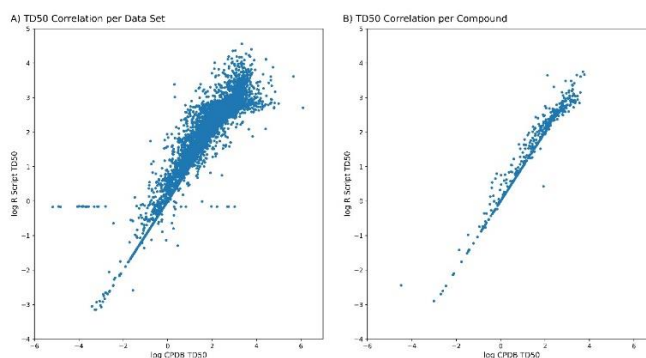

Figure 3: Correlation between log TD_{50} values taken from CPDB and calculated by the R script a) for individual tissue/tumour dose-response data sets, and b) per compound

3.3. Suitability of Data for TD_{50} Modelling

The R script also performs an assessment as to whether the use of a linear model is appropriate for TD_{50} modelling given the dose-response data available (see section 2.3). Using this measure, a total of 1,348 of the 8,956 data points (15.1%) were considered 'unsuitable' for TD_{50} generation via the linear model. Of these 703 (52.2%) generated an outlier value (for our purposes, a log R script TD_{50} greater than 6). This shows a much higher proportion of outliers than the 1,192 (15.7%) data points labelled as outliers in the 7,608 data points considered suitable for the use of a linear model.

Figure 4 shows the correlation between log R script TD_{50} and log CPDB TD_{50} separately for data points assessed as either suitable (figure 4a) or unsuitable (figure 4b) for a linear model. The distribution of the log TD_{50} values from unsuitable data points appears to cluster at a log TD_{50} above 2 (100 mg/kg on the original scale). Removal of these unsuitable data points resulted in a Pearson correlation of 0.934, a slight increase on the 0.926 observed across the whole data set.
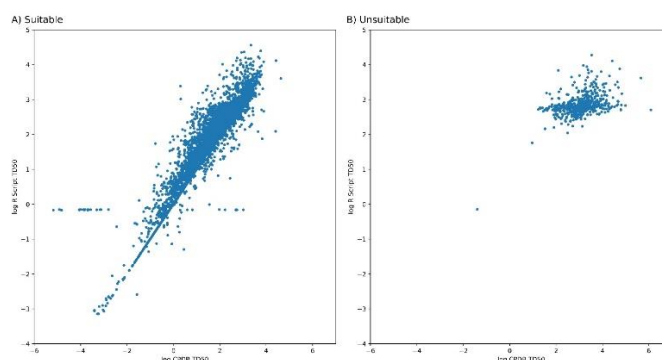

Figure 4: Correlation between log TD50 values taken from CPDB and calculated by the R script where the data set was a) suitable, or b) unsuitable for use in a linear model

ARTICLE

Table 2: Effect of dose range on R script generated TD_{50} values

Data Set	Number of Data Points	Number of Outliers (%) [a]	Number of Unsuitable Data Points (%) [b]	Pearson Coefficient Correlation	
				Without Outliers	Without Outliers and Unsuitable
Low	4,722	417 (8.8)	51 (1.2)	0.923	0.924
Medium	2,119	544 (25.7)	211 (13.4)	0.802	0.866
High	2,115	934 (44.2)	383 (32.4)	0.668	0.753

[a] Number of data points with a log R script TD_{50} >6
[b] After removal of outliers

3.4. Investigating Mismatch Issues

3.4.1. Tested Dose Range

To investigate the relationship between the dose range tested and the subsequent TD_{50} values, the data set was split into three subsets, labelled 'low', 'medium' and 'high'. These consisted of those studies with a highest tested dose below 100 mg/kg ('low'), those with doses ranging from below 100 mg/kg to above 100 mg/kg ('medium'), and those where all tested doses were above 100 mg/kg ('high'). The 100 mg/kg value was chosen as the cut-off point as this appears to be the TD_{50} value beyond which the majority of unsuitable data points were found. Supplementary appendix B shows the correlation of the log R script and log CPDB TD_{50} values for the low, medium and high groups. Table 2 gives the number of data points present in each group, together with the number of outliers, unsuitable data points and Pearson coefficient values. Both the number of TD_{50} outliers and the number of unsuitable data points (both before and after removal of the outliers) increased from the low to medium and high groups, with the high dose group showing the largest proportion of both. At the same time the correlation between the log R script and log CPDB TD50 values, as quantified by the Pearson coefficient, decreased, even after removal of outliers and unsuitable data points. This evidence suggests that there is greater uncertainty inherent in precise TD_{50} values where this equates to doses greater than 100 mg/kg.

3.4.2. Lifetable v summary data

The data downloaded from the CPDB website also included an indication as to which calculation method had been used by the CPDB when deriving the TD_{50} values (that is, the lifetable or summary methods). Separate analysis was made of the relationship between the CPDB and R script TD_{50} values for each of the CPDB calculation methods, as illustrated by Figure 5. The R script method aligned much more closely with the TD_{50} values produced by the summary method than with the lifetable method, showing a Pearson coefficient of 0.989 (Figure 5b), compared to 0.891 (Figure 5a) from the lifetable data set. This confirms the validity of the R script methodology as being a faithful reworking of the principles used in the summary method, which gives
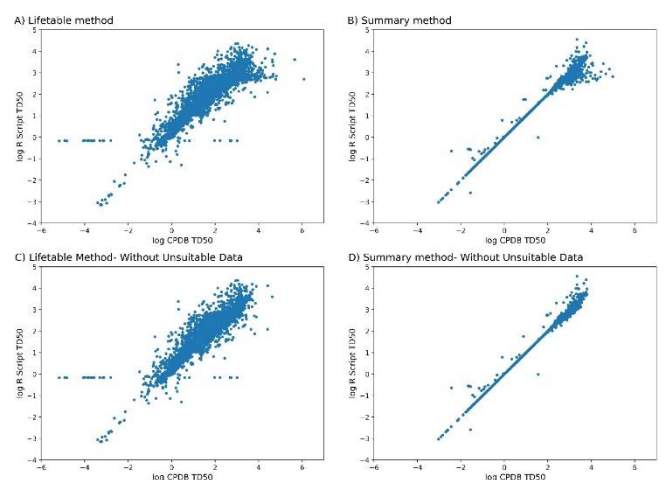

Figure 5: Correlation between log TD_{50} values taken from CPDB and calculated by the R script where CPDB used (A) the lifetable calculation method or (B) the summary calculation method, and the effect of removing unsuitable data points from (C) the lifetable and (D) summary data sets

Table 3: Data set counts and Pearson correlation coefficient after removal of outliers in regard to curve notation grouping

Data Set	Number of Data Points	Number of Outliers (%) [a]	Number of Unsuitable Data Points (%) [b]	Pearson Coefficient Correlation	
				Without Outliers	Without Outliers and Unsuitable
*	6,841	1,702 (24.9)	557 (10.8)	0.929	0.939
\	412	43 (10.4)	33 (8.9)	0.916	0.919
/	652	66 (10.1)	18 (3.1)	0.914	0.926
Z	1,051	84 (8.0)	37 (3.8)	0.920	0.922

[a] Number of data points with a log R script TD_{50} >6
[b] After removal of outliers

confidence in its use to determine TD_{50} values for new data sets.

Figure 5 also shows the effect of removing those data points which were judged as being unsuitable for linear modelling (5c and 5d). 11.8% of the lifetable group were labelled as unsuitable, compared to 4.6% of the summary group. While the proportions of these unsuitable data points are relatively low, their removal had a noticeable effect on the correlation between the log R script and log CPDB TD_{50} values, especially in the summary data set (figure 5d). This removal of unsuitable data points resulted in Pearson coefficients of 0.899 and 0.995 for the lifetable and summary data sets, respectively.

3.4.3. Dose-response group curve

The curve notation provided by the CPDB was used to separately analyse the TD_{50} data generated in relation to the groups described in Table 1. No demonstrable difference in the relationship between log CPDB TD_{50} and log R script TD_{50} values was observed (see supplementary appendix C). The number of data sets, outliers and unsuitable data points, together with the Pearson coefficients for log R script vs log CPDB TD_{50} values are given in Table 3. Given the similarity in Pearson coefficient values, it is unlikely that the assignment of data points to these different groups is a significant factor in the applicability of the TD_{50} model.

In the original CPDB data set there were a number of dose-response sets where the generation of TD_{50} values is questionable. The use of the curve notation highlights this, with category '0' being listed as having either no dose-related effect or containing only a single dose group. However, 5,324 data sets with a '0' curve notation showed a log CPDB TD_{50} of less than 6, indicating that, despite acknowledging these data points as displaying no dose-response relationship, CPDB continued to calculate TD_{50} values from the tumour incidence data.

Similarly, groups \ and Z required the removal of dose points from the tumour incidence data before applying the TD_{50} calculation method. This suggests that the raw data generated by the study was unsuitable for TD_{50} generation and brings the validity of the subsequent TD_{50} values into question. The R script was rerun against the \ and Z curve groups without first applying the pre-calculation processing. The resultant TD_{50} values showed a higher proportion of outliers (23.2%) and unsuitable data points after removal of outliers (14.9%), compared to the same tumour incidence sets run with the pre-processing modifications (8.7% and 5.2%, respectively). Despite this, only a relatively small decrease in the Pearson coefficient correlation between log R script and log CPDB TD_{50} values was observed after removal of the outlier and unsuitable data points (0.891 without modifications, compared to 0.921 with modifications).

This, coupled with the difficulty in replicating the method by which the CPDB curve notation was generated, is problematic when considering the reproducibility of the TD_{50} values.

ARTICLE

Table 4: Data set counts and Pearson correlation coefficient after removal of outliers, with regard to test animal species

Data Set	Number of Data Points	Number of Outliers (%) [a]	Number of Unsuitable Data Points (%) [b]	Pearson Coefficient Correlation	
				Without Outliers	Without Outliers and Unsuitable
Rat	4,797	967 (20.2)	166 (4.3)	0.940	0.941
Mouse	4,032	914 (22.7)	477 (15.3)	0.906	0.920

[a] Number of data points with a log R script TD_{50} >6
[b] After removal of outliers

3.3.4. Species

The correlation between CPDB and R script TD_{50} values was also examined according to the species of test animal used. Analysis was performed on the two most common species; rat and mouse. Other species were present but the number of studies available was too small to allow for meaningful analysis. Information regarding the number of data sets, outliers and unsuitable data points, and the Pearson coefficient (after removal of outliers and of outliers and unsuitable data points) between log TD_{50} values from the CPDB and the R script, are shown in Table 4.

There appeared to be no major difference in correlation between the two species (see supplementary appendix D). Both displayed similar profiles compared to the complete data set, with a greater correlation between the CPDB and R script values for higher potency data sets.

4. Discussion

The results indicate that the TD_{50} values generated using the R script supplied alongside this manuscript correlate with, but do not exactly match, those contained within the CPDB. It is important to acknowledge here that all TD_{50} values are models of the dose-response data, and so there is no single, ideal TD_{50} value for a given data set, there are only values generated through different modelling techniques. That said, the aim was to recreate the methods described by Gold *et al.*, so it was a surprise to find discordance at a level above what might be expected through occasional errors and deviations from protocol.

One of the most influential factors, as demonstrated by Figure 5, is the use of a lifetable model. Questions remain about the exact nature of the lifetable calculation method. The methodology described in the CPDB requires tumour incidence data from a number of time points in order to assess how the probability of tumorigenesis develops throughout the lifespan of the test subjects. However, the standard long-term rodent assay protocol, typically used in carcinogenicity studies, does not provide this information. The data sources given by the CPDB which relate to the lifetable data set predominantly indicate that this data originated with the NTP studies. However, the standard protocol described by the NTP details a long-term treatment ending in a terminal sacrifice without routine interim sacrifices.[18] While interim sacrifices and sacrifice of moribund animals are discussed, these do not form part of the standard carcinogenicity testing protocol. This raises the question of which data were used as part of the lifetable method. Taking tumour incidence data from moribund or spontaneously dead animals would not be sufficient for the requirements of the lifetable method. Another possibility could be that the CPDB group had access to more detailed data than has been subsequently made available in the download files. However, this would cause issues with the reproducibility of the TD_{50} values generated from such data. In any case, the difference in calculations (and input data) explains the differences between TD_{50} values being generated by the R script and the CPDB, where the latter used the lifetable method. The remaining summary data shows a much higher correlation

between the two methods, indicating that the former is fit-for-purpose for the assessment of new data.

Although the effect on discordance was not strong as judged by correlation coefficients, the use of the curve notation described by the CPDB is problematic. This is because the method by which the CPDB assigned a curve to each data set is difficult to reproduce, as it appears to be based on expert judgment. Thus, the assignment of a curve to new data being assessed may be done in an inconsistent manner to those performed previously. More fundamentally, the practice of altering the data sets to be more amenable to TD_{50} calculation is likely to be a larger factor in reproducibility issues for several reasons. The removal of inconvenient data points from a data set in order to produce a clearer dose-response curve may lead to the resultant TD_{50} values not being fully representative of the test compound's tumourigenic potential. The worst-case scenario is where higher doses lead to plateaus of activity at less than 50% – removal of these higher dose data points will give an inaccurate estimate of potency (better, in this case, to use an alternative metric). This will then have the knock-on effect of adversely influencing any metric generated from the TD_{50} values. A more likely scenario within the CPDB is where the removal of some of the top dose data points leads to a reasonable estimate of TD_{50}, but no record is kept of the data points that were removed, again preventing reproducibility. A better alternative may be to recognise those dose-response data sets unsuited to TD_{50} calculation and remove them or use an alternative metric for point of departure – emphasising the requirement for expert review of the data.

The use of the test against a constant relationship in the R script has shown a reproducible way of assessing the suitability of the data for the linear model-based TD_{50} calculation. Data sets designated as unsuitable were shown to have a higher incidence of outliers with a log R script TD_{50} value greater than 6, and a poorer correlation with the published CPDB TD_{50} values. The test also revealed a total of 1,338 data sets that were considered unsuitable for TD_{50} calculation, but which have valid TD_{50} values (that is, log TD_{50} less than 6) as published by the CPDB. A more sophisticated approach for these data points would be to use modelling to generate dose-

response curves that are more amenable to point of departure calculations.

From this analysis it is possible to infer a set of recommendations to assess the applicability of new data for the use of this TD_{50} modelling technique.

- When taking a result from a study published in the literature, data points with no tumorigenic response, as judged by the study authors, should not be considered for TD_{50} modelling. The CPDB included 2,574 data points with a negative authors' call and a log CPDB TD_{50} value less than 6. (section 3.1);
- Studies with one only dose group should not be used for TD_{50} modelling; Bercu et al., (2018)[6] suggested a minimum of three dose groups be present;
- Additional scrutiny should be applied to data sets where the R script described in this paper indicates that the linear model should not be applied (section 3.3) or where the calculated TD_{50} exceeds what would be a reasonable physiological value (section 3.1).
- More generally, when the calculated TD_{50} value exceeds doses of 100 mg/kg, the lower correlation between the methods described here (see section 3.4.1) indicate increased uncertainty in the precise values.

Where these parameters indicate the data as being unfit for TD_{50} modelling an expert review should be undertaken to determine the biological relevance of the dose-response data and suggest possible alternative point-of-departure models.

The difficulties encountered in reproducing previous TD_{50} calculations underline the importance of having documented and hard-coded algorithms for generating metrics which are actively used in human risk assessments. This work has created just that, in an R script that can be widely shared and re-used. A further outcome of this work has been the generation of a new TD_{50} data set. The Lhasa Limited Carcinogenicity Database provides a freely-available, searchable interface through which the data can be accessed.[19] All data downloaded from the CPDB has been transferred to this new system.

Of the TD_{50} values calculated through the R script, all those with a log TD_{50} value greater than 6 have been removed from the published data set. The dose-

response information is still available but no newly calculated TD_{50} values have been displayed in accordance with the above inference that these data sets are not suitable for TD_{50} calculation. It should be noted that no attempt has been made to assess the quality and relevance of the study data, as this would have confounded comparisons with the CPDB. However, the creation of a reproducible methodology for calculating TD_{50} values supports future work for the refinement of, and addition to, the CPDB. The TD_{50} values published in the CPDB remain available in a separate field within the new database, in order to allow easy comparison with the newly generated values.

Conclusions

Overall, an R script has been written and shared in order to provide a reproducible, documented and validated method for generation of TD_{50} values which can be used for risk assessment, benchmarking of alternative metrics or as part of efforts to re-evaluate the TTC for carcinogenicity. This is an aid for point of departure calculations based on dose-response data sets and emphatically does not negate the careful selection of studies or close examination of dose-response data (for human relevance of tumours for example) before generating such metrics.

Conflicts of interest

Dr Gosling reports consultancy fees from Lhasa Limited during the course of the study.

Notes and references

1.   Carcinogenic Potency Project (CPDB). https://toxnet.nlm.nih.gov/cpdb/ (accessed 2nd December 2015).

2.   Chemical Effects in Biological Systems (CEBS). National Toxicology Program (NTP) https://doi.org/10.22427/NTP-DATA-1 (accessed 6th August 2018).

3.   Cheeseman M.A., Machuga E.J., Bailey A.B., A tiered approach to threshold of regulation, *Food Chem. Toxicol.,* 1999, **37**, 387-412. https://doi.org/10.1016/S0278-6915(99)00024-1

4.   Boobis A., el al., Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their re-evaluation, *Crit. Rev. Toxicol.,* 2017, **47**, 710-732. https://doi.org/10.1080/10408444.2017.1318822.

5.   International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk, 2017, **M7**.

6.   Bercu J.P., et al., Potential impurities in drug substances: Compound-specific toxicology limits for 20 synthetic reagents and by-products, and a class-specific toxicology limit for alkyl bromides, *Regul. Toxicol. Pharmacol.,* 2018, **94**, 172-182. https://doi.org/10.1016/j.yrtph.2018.02.001.

7.   Cox D.R., Regression models and life-tables, *J. R. Stat. Soc. Series B Stat. Methodol.* 1972, **34**, 187-220. https://www.jstor.org/stable/2985181.

8.   Peto R., et al., The TD_{50}: A proposed general convention for the numerical description of the carcinogenic potency of chemicals in chronic-exposure animal experiments, *Environ. Health Perspect.,* 1984, **58**, 1-8.

9.   Sawyer C., et al., Calculation of carcinogenic potency from long-term animal carcinogenesis experiments, *Biometrics,* 1984, **40**, 27-40. https://www.jstor.org/stable/2530741.

10. Gold L.S. el al., An empirical comparison of methods used to estimate carcinogenic potency in long-term animal bioassays: Lifetable vs summary incidence data, *Fundam. Appl. Toxicol.,* 1986, **6**, 263-269. https://doi.org/10.1016/0272-0590(86)90239-3.

11. Gold L.S. et al., A carcinogenic potency database of the standardized results of animal bioassays, *Environ. Health Perspect.,* 1984, **58**, 9-319.

12. R Project. The R Project for Statistical Computing, version 3.2.3. https://www.r-project.org/ (accessed 10th December 2015).

13. Chemical Abstract Service (CAS) Registry. https://www.cas.org/support/documentation/chemical-substances (accessed 6th August 2018).

14. Lhasa Limited. Vitic Nexus, Vitic_Lhasa 2016.1 database. https://www.lhasalimited.org/products/vitic.htm (accessed 12th February 2016).

15. Fletcher R., *Practical Methods of Optimization: Vol. 2: Constrained Optimization.* John Wiley & Sons 1981.

16. KNIME Analytics Platform, version 3.1.2. https://www.knime.com/ (accessed 4th January 2016).

17. Python Programming Language, version 3.6.4. https://www.python.org/ (accessed 19th December 2017).

18. National Toxicology Program (NTP), Specifications for the conduct of studies to evaluate the toxic and carcinogenic potential of chemical, biological and physical agents in laboratory animals for the National Toxicology Program (NTP),

2011. https://ntp.niehs.nih.gov/ntp/test_info/finalntp_toxcarspecsjan2011.pdf.

19. Lhasa Limited. Carcinogenicity Database. https://carcdb.lhasalimited.org/ (accessed 6th August 2018).