Webinar: Accessing Carcinogenicity Potency Data

1st April 2020

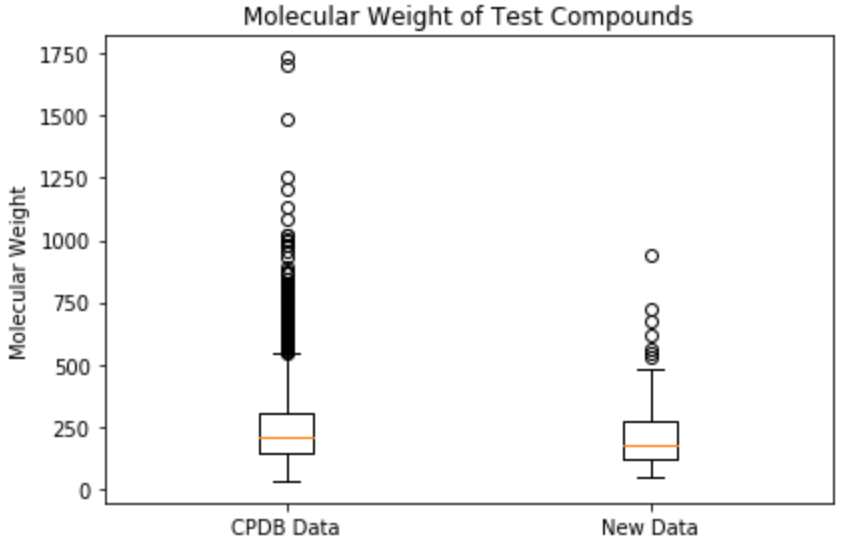
Andrew Thresher

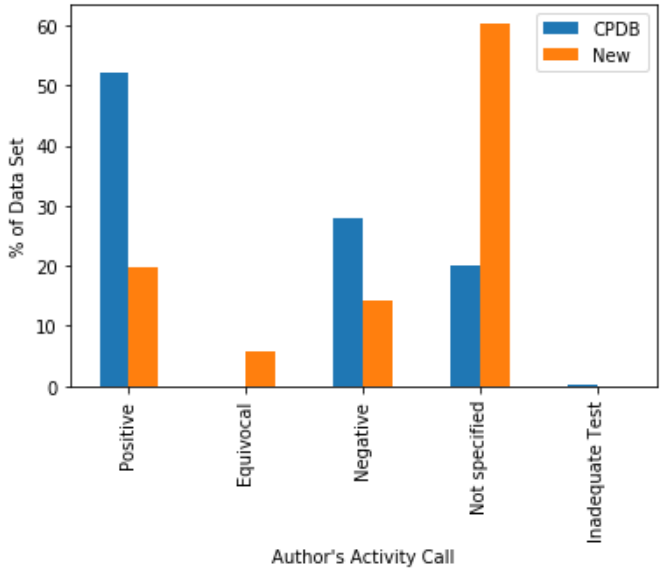
Data

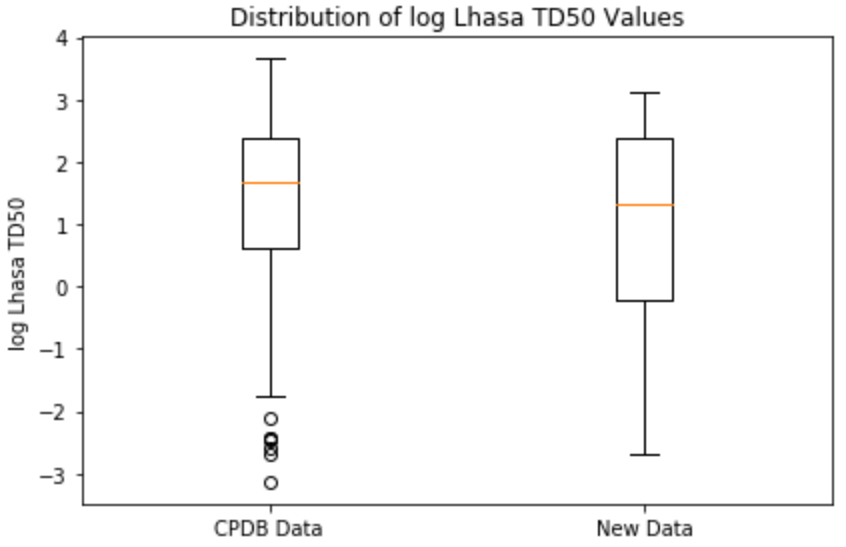
- The data from CPDB has been faithfully reproduced in the Lhasa Carcinogenicity DB.
  - 1,529 test compounds,
  - 6,529 long-term carcinogenicity study records.

- Additional data gathered from National Toxicology Program (NTP)
  - 197 additional test compounds,
  - 1,216 additional study records.

New Data


Molecular Weight of Test Compounds
Molecular Weight
1750
1500
1250
1000
750
500
250
0
CPDB Data
New Data


CPDB
New
% of Data Set
60
50
40
30
20
10
0
Positive
Equivocal
Negative
Not specified
Inadequate Test
Author's Activity Call


Distribution of log Lhasa TD50 Values
log Lhasa TD50
4
3
2
1
0
−1
−2
−3
CPDB Data
New Data

CPDB

Formaldehyde (CAS 50-00-0)

SMILES, InChI and Structure are below.

Rats and Mice: Cancer Test Summary

Rat Target Sites		Mouse Target Sites		TD_{50} (mg/kg/day)	
Male	Female	Male	Female	Rat	Mouse
hmo nas	hmo nas	nas	no positive	1.35[m,v]	43.9

Hamsters: Cancer Test Summary

Hamster Target Sites		TD_{50} (mg/kg/day)
Male	Female	
no positive	no test	no positive

Chemical (Synonym) CAS
Species Sex Strain Route Xpo+Xpt PaperNum 0 Dose 1 Dose 2 Dose 3 Dose Literature Reference or NCI/NTP:Site Path
Site Path Notes TD50 DR Pval AuOp LoConf UpConf Cntrl 1 Inc 2 Inc 3 Inc Brkly Code

FORMALDEHYDE 50-00-0

2890	H	m	syg	inh	94w94	1414m		0	.772mg			Dalbey;txcy,24,9-14;1982
res	tum	r	no	dre	P=1.	-	6.50mg	n.s.s.	0/50	0/50		
2891	H	m	syg	inh	25m25	1414n		0	1.29mg			
res	tum	rs	no	dre	P=1.	-	25.6mg	n.s.s.	0/132	0/88		
2892	M	f	b6c	inh	52w52	1566a		0	.772mg	2.32mg	5.79mg	Pavkov;ciit;1981/Kerns 1983
nas	tum	ek	no	dre	P=1.		.271mg	n.s.s.	0/10	0/10	0/10	0/10
lun	tum	ek	no	dre	P=1.		2.98mg	n.s.s.	0/10			0/10
liv	tum	ek	no	dre	P=1.		2.98mg	n.s.s.	0/10			0/10
2893	M	f	b6c	inh	78w78	1566b		0	.772mg	2.32mg	5.79mg	
nas	tum	ek	no	dre	P=1.		1.21mg	n.s.s.	0/20	0/20	0/20	0/19
liv	mix	ek	no	dre	P=1.		8.42mg	n.s.s.	2/20			1/19
lun	tum	ek	no	dre	P=1.		12.8mg	n.s.s.	0/20			0/19
2894	M	f	b6c	inh	24m24	1566c		0	5.79mg			
lun	mix	ek	47.3mg		P<.3		10.9mg	n.s.s.	1/30	3/27		
liv	mix	ek	98.7mg		P<.6		14.4mg	n.s.s.	1/30	2/28		
nas	tum	ek	no	dre	P=1.		3.10mg	n.s.s.	0/30	0/26	0/41	0/28
2895	M	f	b6c	inh	24m27	1566m		0	.686mg	2.06mg	5.15mg	
nas	tum	e	no	dre	P=1.		6.06mg	n.s.s.	0/50	0/54	0/39	0/54
lun	mix	e	no	dre	P=1.		44.5mg	n.s.s.	1/50			1/54
liv	tum	e	no	dre	P=1.		72.5mg	n.s.s.	0/50			0/54
2896	M	m	b6c	inh	52w52	1566a		0	4.83mg			
liv	hpc	ek	7.85mg		P<.3		1.28mg	n.s.s.	0/10	1/10		
lun	tum	ek	no	dre	P=1.		2.49mg	n.s.s.	0/10	0/10		
nas	tum	ek	no	dre	P=1.		.226mg	n.s.s.	0/10	0/10	0/10	0/10
2897	M	m	b6c	inh	24m24	1566c		0	.644mg	1.93mg	4.83mg	
ntu	sqc	es	43.9mg	*	P<.04	+	10.8mg	n.s.s.	0/20	0/22	0/19	2/17
liv	mix	es	34.5mg		P<.2		9.67mg	n.s.s.	4/62			6/40
lun	mix	es	no	dre	P=1.		6.53mg	n.s.s.	6/25			4/18

Lhasa Carcinogenicity DB

Formaldehyde

Summary

Species	Lhasa TD_{50} (mg/kg/day)	Gold TD_{50} (mg/kg/day)	Result	Sex	Tumour sites
Hamster	-	-	NEGATIVE	♂ Male	-
Mouse	44.3	43.9	NO POSITIVE EVIDENCE	♀ Female	-
			POSITIVE	♂ Male	Nasal cavity- turbinate
Rat	1.67	1.35	POSITIVE	♀ Female	Gastrointestinal tract, Multiple tumour sites, Nasal cavity- turbinate
			POSITIVE	♂ Male	Gastrointestinal tract, Multiple tumour sites, Nasal cavity, Nasal cavity- mucosa, Nasal cavity- turbinate

Study details and citations

expand all | collapse all

Species: Mouse | Sex: Male | Strain: B6C3F1 | Route: Inhalation | Exposure time: 104 week(s) | Experiment time: 104 week(s)

Tumour Site	Tumour Type	Probability		Lhasa TD_{50}	Gold TD_{50}
Liver	Multiple tumour types	<= 0.162	NOT SPECIFIED	-	34.5
Lung	Multiple tumour types	= 1	NOT SPECIFIED	-	-
Nasal cavity- turbinate	Carcinoma- squamous cell	<= 0.039	POSITIVE	44.3	43.9

Dose (ug/kg/day)	0	644	1930	4830
Incidence	0/20	0/22	0/19	2/17

Literature reference(s)

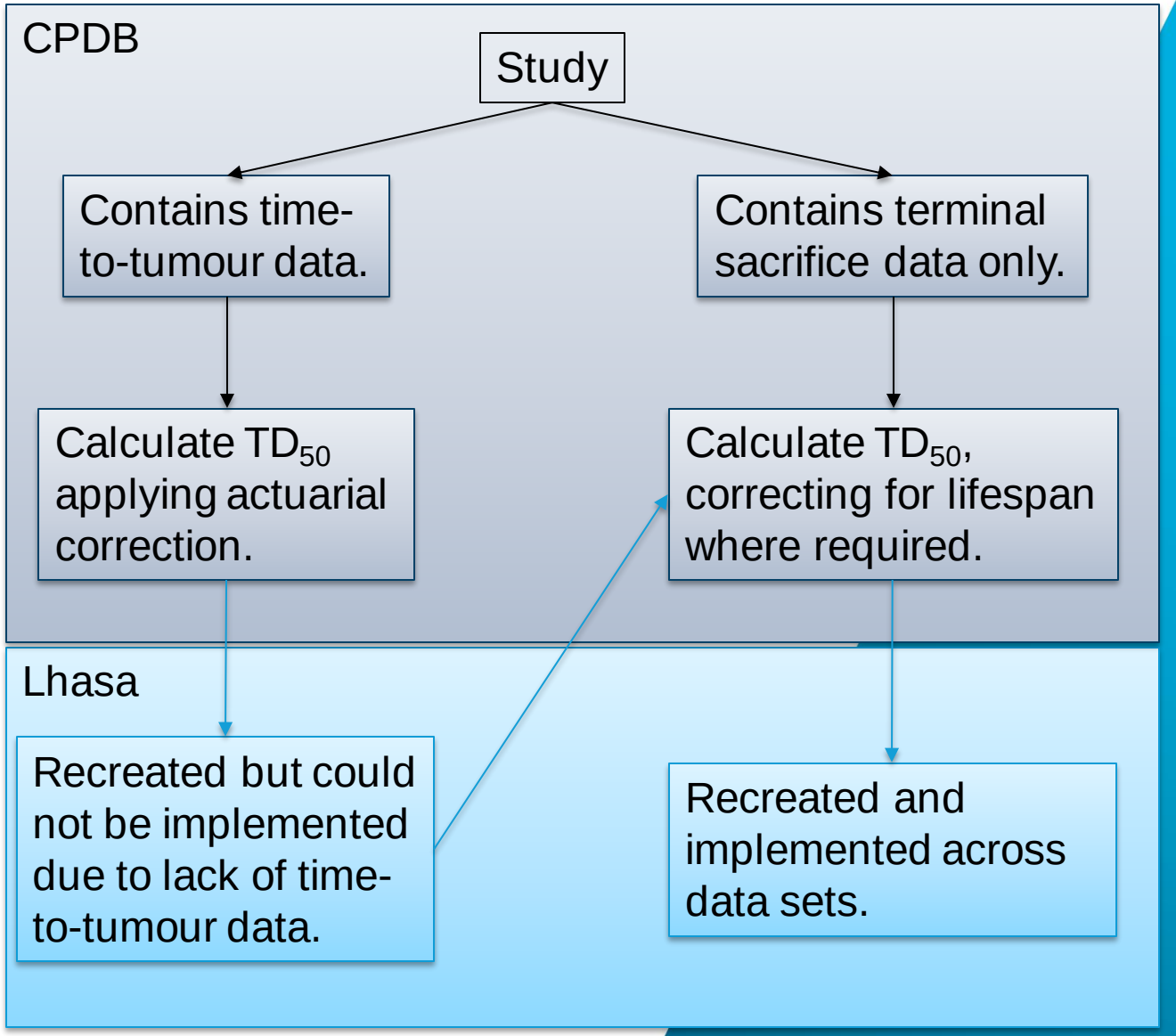
Author	Title	Journal	Year	Volume	Page	Reference type
Kerns WD, Pavkov KL, Donofrio DJ, Gralla EJ, Swenberg JA	Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure	Cancer Research	1983	43	4382-4392	Additional reference
Pavkov KL, Kerns WD, Mitchell RI, Connell MM, Donofrio DJ, Harroff HH, Barker AD, Fisher GL	A chronic inhalation toxicology study in rats and mice exposed to formaldehyde		1981			Primary reference

Notes (mortality)

Mortality
Test subject survival was decreased

TD_{50} Calculation

- Reproduce the CPDB calculation method described in:
  - Cox D.R., (1972) J. R. Stat. Soc. Series B Stat. Methodol. 34, 187-220.
  - Peto R. et.al., (1984) Environ. Health Perspect. 58, 1-8.
  - Sawyer C. et.al., (1984) Biometrics 40, 27-40.
- Work done in partnership with the University of Leeds.



Implementation

- CPDB assigned each tumour incidence a ‘curve’ to denote the linearity of the dose-response. This was used to remove data points in order to isolate the linear portion of the dose-response.

- We were unable to reproduce the mechanism by which CPDB assigned the curve notation.
  - For the CPDB data we used their assigned curve notation to indicate which data points needed to be removed,
  - For new data we did not remove any data points,
  - We included a measure of whether the dose-response curve aligns more with a linear model or a constant model, and so whether it is suitable for use in the TD_{50} model.

- Our method was published in Toxicology Research (2019) 8, 696-703.

ARTICLE

Generation of TD_{50} Values for Carcinogenicity Study Data

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

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

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.

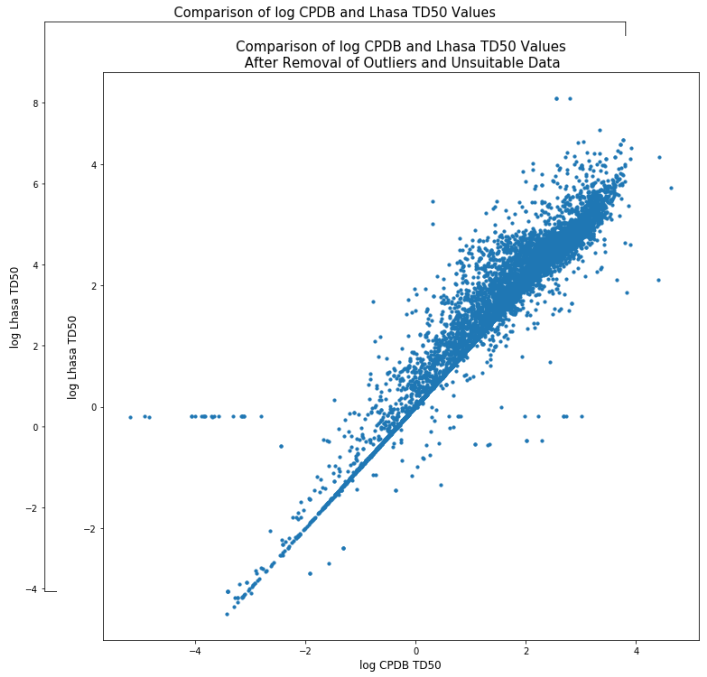
Historically, the metric used to determine carcinogenic potency has been the TD_{50}, defined as the dose required to halve the probability of a subject remaining without tumours throughout a lifetime of exposure. TD_{50} values were included in the CPDB and subsequently used to derive the threshold 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
Electronic Supplementary Information (ESI) available: [Supplementary appendix A- the R script for generating TD50 values; B- analysis by tested dose range; C- analysis by CPDB curve notation; D- analysis by test subject species]. See DOI: 10.1039/x0xx00000x

Comparison of CPDB and Lhasa TD_{50} Values

- 8.6% did not produce a Lhasa TD_{50}
- Initial comparisons show a broadly good correlation, with collections of data points showing excessively high log TD_{50} values :
  - Low tumour incidence, even at high doses,
  - Low number of animals
  - Curve notation suggests no dose-related effect
- After removal of these outliers and data points considered unsuitable by the Lhasa model, the Pearson correlation coefficient = 0.952.

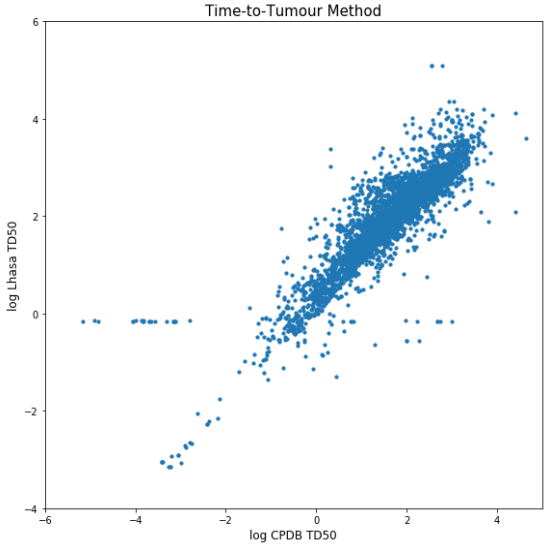


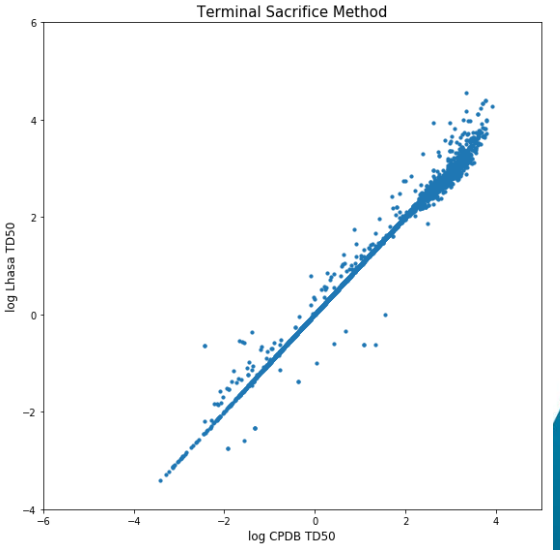
Dose (mg/kg/day)	0	305	406	569	731
Incidence	0/30	1/30	1/30	2/30	2/30

Time-to-Tumour vs Terminal Sacrifice Methods

- When the data is separated according to the method used by CPDB, the terminal sacrifice method shows a much closer correlation with the Lhasa TD_{50} values.

- This is expected, as the Lhasa model is based on the terminal sacrifice model, but this highlights the issues of reproducibility arising from the lack of time-to-tumour data.

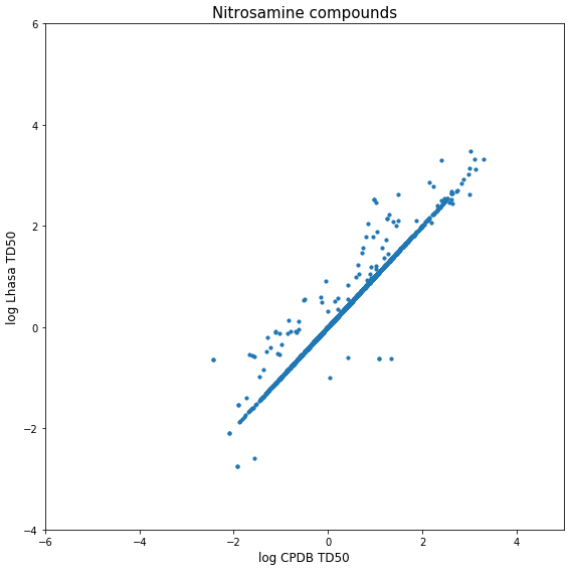


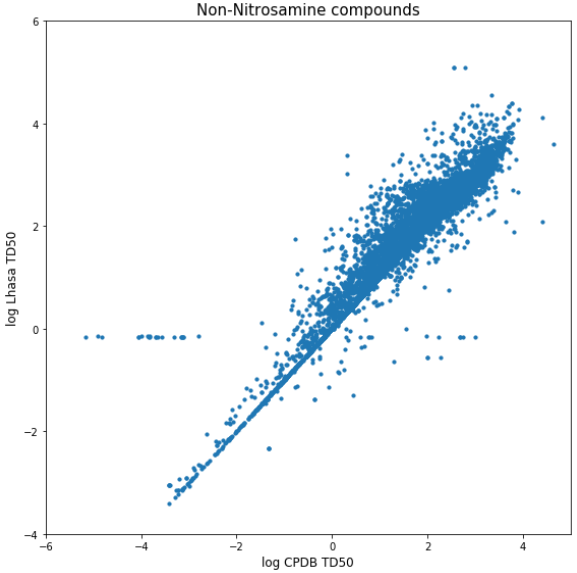


Nitrosamine Data

- The data set contained 139 nitrosamine compounds.
  - 119 compounds with positive results
  - All produced Lhasa TD_{50} values with 1.9% showing outlying values.

- Pearson coefficient:
  - Nitrosamines: 0.973
  - Non-nitrosamines: 0.943

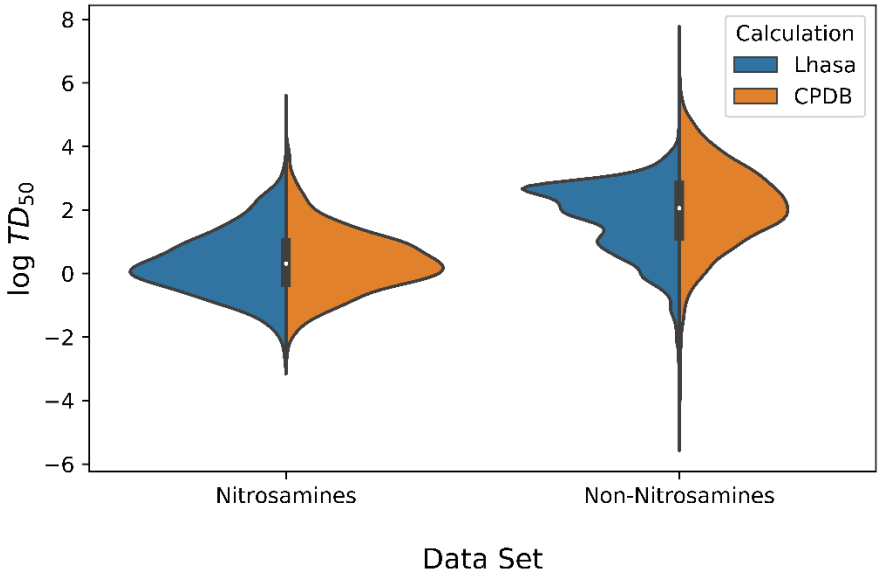




Nitrosamine Data

- The nitrosamine data shows a more potent TD_{50} distribution than the non-nitrosamine compounds.

- Median log Lhasa TD_{50} values:
  - Nitrosamines – -0.33
  - Non-nitrosamines – 1.65



Differences between CPDB and Lhasa TD_{50} values

- Given that the TD_{50} values are being generated from the same dose-response data and attempting to use the same methodology, we would expect the two sets of values to be identical. While some data points do align very closely, other show more variation.
  - Time-to-tumour vs terminal sacrifice method,
  - Curve notation and pre-processing data to make it more amenable to modelling,
  - Use of unsuitable dose-response data.

Differences between CPDB and Lhasa TD_{50} values

- No objective TD_{50} value. Rather these are models of the data and so different techniques will produce different values.
- Good correlation with previous publications provides confidence that the method can be applied to new data.
- The transparency of our model means TD_{50} calculation can be challenged and reproduced.
- Critical evaluation of the input data is also important:
  - Is it suitable for dose-response modelling?
  - Is it reasonable to extrapolate to a 50% response?