# Predicting Chronic Non-Cancer Toxicity Levels from Short-Term Toxicity Data

by Jessica Kratchman

M.S. in Fire Protection Engineering, May 2007, Univ. of Maryland, College Park
B.S. in Fire Protection Engineering, Dec. 2004, Univ. of Maryland, College Park

A Dissertation submitted to

The Faculty of
The Milken Institute School of Public Health
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Public Health

May 21, 2017

George Gray
Professor of Environmental and Occupational Health

The Milken Institute School of Public Health of The George Washington University certifies that Jessica Kratchman has passed the Final Examination for the degree of Doctor of Public Health as of March 15, 2017. This is the final and approved form of the dissertation.

Predicting Chronic Non-Cancer Toxicity Levels from Short-Term Toxicity Data

Jessica Kratchman

Dissertation Research Committee:

George Gray, Professor of Environmental and Occupational Health, Dissertation Director

John Fox, Committee Member

Bing Wang, Professor of Food Science and Technology, Committee Member

Matias Attene Ramos, Professor of Environmental and Occupational Health, Committee Member

# Acknowledgments

I would like to acknowledge the input and support I received from my dissertation committee and my outside readers Dr. Carol Henry and Dr. Richard Engler. I appreciate their time and expertise tremendously. I would also like to acknowledge my husband and daughter, Dennis and Emily Andrukat, for supporting me throughout my degree. My husband provided unparalleled commitment and his support provided me the ability to juggle it all. Also for my daughter, I completed this for you because as we say, *there is no such thing as a girl's job*, *girl power*! I would also like to thank my parents, brother and late grandparents. Thanks for inspiring me and providing me with the tools, curiosity and drive to be a scientist.

# Abstract of Dissertation

## Predicting Chronic Non-Cancer Toxicity Levels from Short-Term Toxicity Data

This dissertation includes three separate but related studies performed in partial fulfillment of the requirements for the degree of Doctor of Public Health in Environmental and Occupational Health. The main goal this dissertation was to develop and assess quantitative relationships for predicting doses associated with chronic non-cancer toxicity levels in situations where there is an absence of chronic toxicity data, and to consider the applications of these findings to chemical substitution decisions. Data from National Toxicology Program (NTP) Technical Reports (TRs) (and where applicable Toxicity Reports), which detail the results of both short-term and chronic rodent toxicity tests, have been extracted and modeled using the Environmental Protection Agency's (EPA's) Benchmark Dose Software (BMDS). Best-fit minimum benchmark doses (BMDs) and benchmark dose lower limits (BMDL) were determined. Endpoints of interest included non-neoplastic lesions, final mean body weights and mean organ weights. All endpoints were identified by NTP Pathologists in the abstract of the TRs as either statistically or biologically significant. A total of 41 chemicals tested between 2000 and 2012 were included with over 1700 endpoints for short-term (13 week) and chronic (2 year) exposures.

Non-cancer endpoints were the focus of this research. Chronic rodent bioassays have been used by many methodologies in predicting the carcinogenic potential of chemicals in humans (1). However, there appears to be less emphasis on non-cancer endpoints. Further, it has been shown in the literature that there is little concordance in cancerous endpoints between humans and rodents (2). The first study, *Quantitative*

*Relationship of Non-Cancer Benchmark Doses in Short-Term and Chronic Rodent Bioassays* (Chapter 2), investigated quantitative relationships between non-cancer chronic and short-term toxicity levels using best-fit modeling results and orthogonal regression techniques. The findings indicate that short-term toxicity studies reasonably provide a quantitative estimate of minimum (and median) chronic non-cancer BMDs and BMDLs.

The next study, *Assessing Implicit* Assumptions in Toxicity Testing Guidelines (Chapter 3) assessed the most sensitive species and species-sex combinations associated with the best-fit minimum BMDL10 for the 41 chemicals. The findings indicate that species and species-sex sensitivity for this group of chemicals is not uniform and that rats are significantly more sensitive than mice for non-cancerous outcomes. There are also indications that male rats may be more than the other species sex groups in certain instances.

The third and final study, *Comparing Human* Health Toxicity of Alternative Chemicals (Chapter 4), considered two pairs of target and alternative chemicals. A *target* is the chemical of concern and the *alternative* is the suggested substitution. The alternative chemical lacked chronic toxicity data, whereas the target had well studied non-cancer health effects. Using the quantitative relationships established in Chapter 2, *Quantitative Relationship of Non-Cancer Benchmark Doses in Short-Term and Chronic Rodent Bioassays,* chronic health effect levels were predicted for the alternative chemicals and compared to known points of departure (PODs) for the targets. The findings indicate some alternatives can lead to chemical exposures potentially more toxic than the target chemical.

Table of Contents

Acknowledgments iv
Abstract of Dissertation v
Predicting Chronic Non-Cancer Toxicity Levels from Short-Term Toxicity Data v
Table of Contents vii
List of Figures ix
List of Tables x
Chapter 1. Introduction 1
Chapter 2. Quantitative Relationship of Non-Cancer Benchmark Doses in Short-Term and Chronic Rodent Bioassays 6
1 Introduction 6
2 Methods 9
2.1 Data collection 9
2.2 Batch-processing and modeling in BMDS 11
2.3 Filters 12
2.4 Relation between chronic and short-term BMDs and BMDLs 15
3 Results 19
3.1 Chemical inclusion results 19
3.2 Best-fit, minimum and median BMD and BMDLs 20
3.3 Results of orthogonal regressions 27
4 Discussion 33
4.1 Study limitations 35
4.2 Conclusions 36
Chapter 3. Assessing Implicit Assumptions in Toxicity Testing Guidelines 38
1 Background 38
1.1 Guidelines for Species-Sex Selection in Toxicity Selections 39
2 Methods 40
2.1 Data Collection 41
2.2 Batch-Processing and Modeling in BMDS 43
2.2.1 Filters 45
2.3 Statistical Analysis 47
3 Results 48
3.1 Chemical Inclusion Results 49
3.2 Minimum, Best-Fit BMDL per Chemical-Duration 50
3.3 Minimum, Best-Fit BMDL Distribution by Species-Sex 53

4 Discussion ........................................................................................................ 57
4.1 Study Limitations ........................................................................................ 57
4.2 Conclusions .................................................................................................. 58
Chapter 4. Comparing Human Health Toxicity of Alternative Chemicals ................. 60
1 Introduction ........................................................................................................ 60
2 Background ........................................................................................................ 64
3 Methods ............................................................................................................. 66
3.1 Identification of Alternative-Target Chemical Study Subjects and PODs ......... 67
3.2 Alternative chemical: Prediction of Chronic Oral BMDL10 ............................ 71
3.3 Comparing Target-Alternative Chronic Oral BMDL10s .................................. 73
4 Results ............................................................................................................... 74
4.1 Assessment of alternatives ............................................................................ 74
5 Discussion ........................................................................................................ 79
5.1 Study limitations ........................................................................................... 80
5.2 Conclusions .................................................................................................. 81
Chapter 5. Conclusions ........................................................................................... 82
References ............................................................................................................... 85

## List of Figures

Figure 1: Generalized approach to chemical assessment.................................................... 2
Figure 2: Minimums: best-fit, BMD and BMDL scatter plots (Log10) ........................... 28
Figure 3: Median results: Best-fit, BMD and BMDL scatter plot results (Log10)........... 30
Figure 4: Least-squares regression minimum BMDL10 .................................................. 32
Figure 5 Target-alternative pair identification.................................................................. 68
Figure 6: Alternative chemical: predicting minimum best-fit chronic BMDL10............. 71
Figure 7: Transfer point 2, comparing chronic BMDL10s of target-alternative pair ....... 74

## List of Tables

Table 1: Sample log diagnostic results: minimum, best-fit BMDLs. .............................. 16
Table 2: Sample log diagnostic results: median, best-fit BMDLs. .................................. 17
Table 3: NTP analyzed chemicals meeting chemical inclusion criteria. .......................... 19
Table 4: Best-fit, minimum BMD10 and BMDL10 absolute results............................... 21
Table 5: Best-fit, minimum BMD50 and BMDL50 results. ........................................... 22
Table 6: Best-fit, median BMD10 and BMDL10 dichotomous datasets. ........................ 24
Table 7: Best-fit, median BMD10 and BMDL10, continuous datasets ........................... 25
Table 8: Best-fit, median BMD50 and BMDL50 results. ................................................ 26
Table 9: Orthogonal regression analysis statistical assessments ..................................... 33
Table 10: NTP analyzed chemicals meeting chemical inclusion criteria. ........................ 49
Table 11: Absolute minimum, Best-fit BMDLs per chemical-duration. .......................... 51
Table 12: Total number of datasets before and after the application of filters. ................ 53
Table 13: Absolute minimum best-fit BMDL goodness-of-fit results, species-sex ......... 54
Table 14: Dichotomous minimum best-fit BMDL goodness-of-fit, species-sex ............. 54
Table 15: Continuous minimum best-fit BMDL goodness-of-fit, species-sex ................ 54
Table 16: Absolute minimum best-fit BMDL pairwise comparison, species-sex ........... 55
Table 17: Dichotomous minimum best-fit BMDL pairwise comparison, species-sex ..... 55
Table 18: Continuous minimum best-fit BMDL pairwise comparison species-sex ......... 56
Table 19: Absolute minimum best-fit BMDL pairwise comparison, species .................. 56
Table 20: Dichotomous minimum best-fit BMDL pairwise comparison, species ........... 56
Table 21: Continuous minimum best-fit BMDL pairwise comparison, species .............. 57
Table 22: MTBE short-term data. ..................................................................................... 76
Table 23: d-Limonene short-term data. ............................................................................ 77
Table 24: Results of Transfer Points 1 and 2; comparison of target-alternative pairs...... 79

# Chapter 1. Introduction

Estimates of relative chemical toxicity are often used to evaluate the magnitude and type of hazards to humans and the environment from chemical contaminants (3). Single chemical assessments and alternatives assessments are performed with increasing frequency due to public demand for "safer" chemicals (4). These tools are often used to characterize threats to human health and identify possible chemical alternatives that reduce the hazard level posed from a particular chemical (5). However, it is often unclear what "safer" even means or if the chemicals being compared are comparable.

The U.S. Environmental Protection Agency (EPA) recommends that when assessing chemicals for chronic exposure that attention be paid to the *critical effect* in the *critical study,* which is the most sensitive adverse effect in animal bioassays "that occurs in the most sensitive species as the dose rate of an agent increases" (6). However, hundreds of new chemicals are developed each year (7) and existing chemicals have different levels of toxicity data available. Therefore many of these new or poorly tested chemicals do not have chronic toxicity data and the chronic critical effects have not been assessed. To address this problem, this dissertation research investigated the use of short-term non-cancer toxicity data as a reasonable predictor of the doses associated with chronic non-cancer toxicity levels. Non-cancer endpoints were the focus of this research. Chronic rodent bioassays have been used by many methodologies in predicting the carcinogenic potential of chemicals in humans (1). However, there appears to be less emphasis on non-cancer endpoints.

This dissertation research addressed three separate but related aspects of relative chemical toxicity levels where data imbalances exist. *Safety*, in the context of this research always refers to human health toxicity levels (i.e. doses). Each study (also referred to as *papers*) presented in this dissertation builds upon the knowledge gained from the previous paper and advances through a general framework for assessing relative chemical safety, presented in Figure 1. Figure 1 is a three step flowchart of standard processes observed within chemical assessment frameworks for establishing and comparing human health reference values. A standalone research paper is associated with each of three major steps, and links the overall research of this dissertation with the goal of making human health determinations in assessing relative toxicity of chemicals when data imbalances exist.

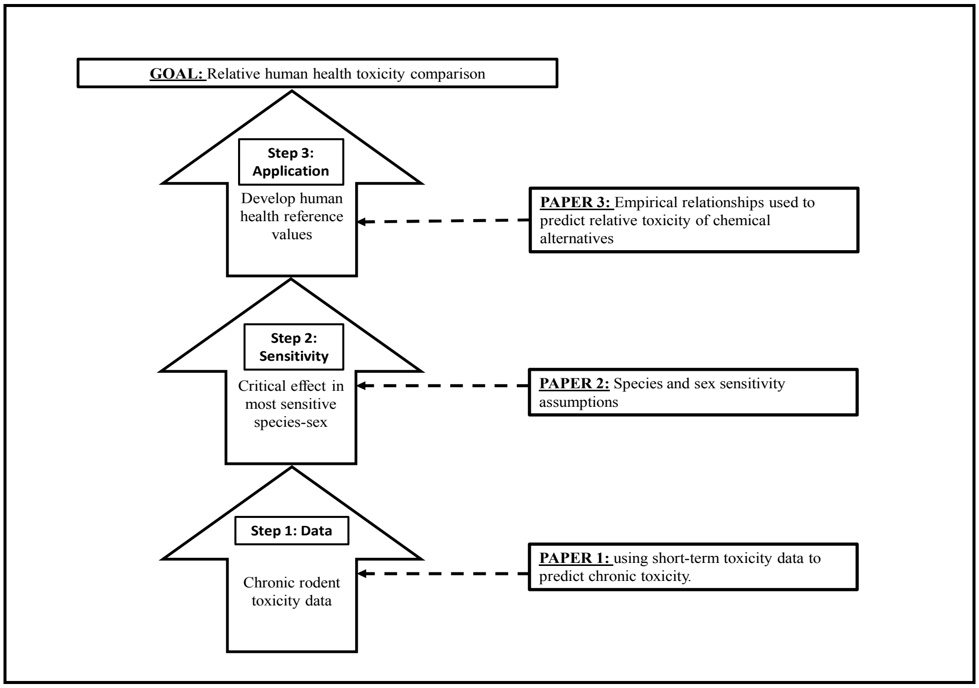

GOAL: Relative human health toxicity comparison
Step 3: Application
Develop human health reference values
PAPER 3: Empirical relationships used to predict relative toxicity of chemical alternatives
Step 2: Sensitivity
Critical effect in most sensitive species-sex
PAPER 2: Species and sex sensitivity assumptions
Step 1: Data
Chronic rodent toxicity data
PAPER 1: using short-term toxicity data to predict chronic toxicity.

Figure 1: Generalized approach to chemical assessment

Paper 1, *Quantitative Relationship of Non-Cancer Benchmark Doses in Short-Term and Chronic Rodent Bioassays* (Chapter 2), addresses Step 1 of Figure 1, the acquisition of chronic rodent toxicity data. The goal of this study was to investigate the quantitative relationship between short-term and chronic non-cancer toxicity levels. In the absence of the necessary chronic non-cancer toxicity data, this study proposes that short-term non-cancer toxicity data may be able to reasonably predict chronic toxicity levels by focusing on the dose-response relationship instead of a critical effect.

Data from National Toxicology Program (NTP) technical reports were extracted and modeled using the EPA's Benchmark Dose Software (BMDS). Using NTP Technical Reports (TRs) (and where necessary Toxicity reports), best-fit minimum benchmark dose (BMD) and benchmark dose lower limits (BMDLs) were modeled for pathologist identified significant non-neoplastic lesions, final mean body weight and mean organ weight of 41 chemicals tested by NTP between 2000 and 2012. Dichotomous and continuous data were considered in this study because both are sensitive toxicity markers which regulatory bodies such as the EPA to set reference values (8).

Models were then developed at the chemical-level using orthogonal regression techniques to predict chronic (2-years) non-cancer health effect levels using the results of the short-term (3-months) toxicity data. The goal in determining these relationships was to identify methods for the faster development of human health toxicity values the assessment for chemicals that lack chronic toxicity data.

The next study, Paper 2: *Assessing Implicit* Assumptions in Toxicity Testing Guidelines (Chapter 3), addresses the second step of the chemical assessment process presented in Figure 1, evaluating the most sensitive species, sex and endpoint. A common goal of toxicity testing is identification of sensitive adverse endpoints for the most sensitive species and sex, also called a *critical effect* or *study* that occur at the lowest exposure levels (9). Traditional toxicology testing guidelines recommend using mice and rats of both sexes as test subjects to predict human health toxicity outcomes (2) (10) (11) (12).

This recommendation appears to be intended to maintain consistency and comparability between tests and testing protocols.

An operationalized approach to assessing this recommendation is that, prior to testing, the predicted distribution of species and species-sex sensitivity over a large group of chemicals is a uniform distribution. This study refers to this approach as an *implicit assumption of uniform species-sex sensitivities*, and studied it by estimating non-cancer BMDL10s for the 41 NTP chemicals, both chronic and short-term exposures, and then assessing the species-sex distribution of the most sensitive groups.

The third and final study, *Comparing Human* Health Toxicity of Alternative Chemicals (Chapter 4), assessed the final step of the process presented in Figure 1, which is the application and practical use of chemical risk information. This paper considered two pairs of target and alternative chemicals proposed by an actual chemical substitution assessment tool. This tool was deemed prototypical of alternative assessment tools available to the public. The results of this study are not intended to be critical of any particular tool, but an assessment of chemical alternative assessment in general.

The alternative chemicals chosen lacked chronic toxicity data, whereas the targets each have well studied non-cancer health effect levels. Using the quantitative relationships established in Chapter 2, *Quantitative Relationship of Non-Cancer Benchmark Doses in Short-Term and Chronic Rodent* Bioassays*,* chronic health effect levels were predicted for the alternative chemicals and compared to known points of departure (PODs) for the targets. The goal was to shed light on possible inconsistencies in comparing chemicals and highlight the need for quantitative predictive techniques when chronic bioassays are not available (13).

The three papers outlined in this section are presented in the next three chapters as stand-alone studies.  Provided in each paper is the background of the issue, the methodology used, the results obtained, a discussion of the findings, and study conclusions and limitations. The final chapter of this dissertation presents the overall conclusions from this research project.

# Chapter 2. Quantitative Relationship of Non-Cancer Benchmark Doses in Short-Term and Chronic Rodent Bioassays

## 1 Introduction

Chronic toxicity tests are frequently the basis for chronic non-cancer human health reference values HHRVs (13). However, there is a notable lack of available chronic toxicity data necessary to derive benchmarks for chemicals and properly assess their potential human health effect levels (7) (13). In the absence of the chronic non-cancer toxicity data, this study investigated whether short-term non-cancer toxicity data can be used to predict chronic health effect levels by focusing on the dose-response relationship instead of specific health endpoints. Non-cancer endpoints were the focus of this research. Chronic rodent bioassays have been used by many methodologies in predicting the carcinogenic potential of chemicals in humans (1). However, there appears to be less emphasis on non-cancer endpoints.

This study utilized chronic (2-year) and short-term (3-month) non-cancer toxicity data from 41 chemicals studied by NTP and published in TRs (and where necessary Toxicity reports for more detail). TRs were included in this study only if they used rat and mice models and had oral exposure routes (feed, water or gavage). This approach does not require concordance of the outcomes, which have been proven to be weak (2), but instead focuses on the exposure or dose required to elicit some adverse non-cancer health endpoint without specification of the effect.

The EPA recommends that when identifying chronic non-cancer risk that "particular attention [be paid]… to the *critical effect*" (6). The *critical effect* in the *critical study* is the most sensitive adverse effect in chronic animal bioassays "that occurs in the most sensitive species as the dose rate of an agent increases" (6). However, obtaining chronic animal bioassays has practical limitations, requiring tremendous time and resource expenditures to obtain the bioassays (7). Additionally, hundreds of new chemicals are being synthesized every year and "innovative approaches need to be developed to determine their health effects in addition to those chemicals that currently lack toxicity data" (7).

A variety of approaches exist for characterizing HHRVs for chemicals that lack chronic non-cancer health data (14). *Toxicity Testing in the 21st Century* (Tox21), and *ToxCast*, are U.S. federal government approaches to developing methods to characterize or improve models for non-cancer health effects from chemicals without chronic animal bioassays using high throughput models. They are intended for use in prioritizing animal testing or for use in chemical risk assessments that lack the available chronic toxicity data (15). The work of Pennington, *et al*. (16) proposes techniques for the evaluation and screening of non-cancer toxicology effects by deriving $\beta_{\mathrm{ED10}}$ slope factors from bioassay and determining possibilities for "extrapolation from other more readily available measures" (16) such as no observed or lowest observed adverse effect levels (NOAELs and LOAELs). Another approach is the development of predictive models based on the quantitative structure-activity relationship (QSAR) methodology. This approach assumes there is a correlation between molecular and structural parameters and bioactivity within a group of chemicals (17).

This study proposes that, in the absence of the necessary chronic non-cancer toxicity data, by focusing on the dose-response relationship instead of a critical effect, short-term data can be used to predict chronic non-cancer health effect levels. This is a novel approach. Based on a review of publically available chemical substitution and single chemical assessment frameworks there is typically a focus on hazard instead of dose or effect level. This study extracted chronic and short-term data from 41 NTP TRs pertaining to non-neoplastic lesions, final mean BW and mean OW tested by NTP between 2000 and 2012. Dichotomous and continuous data were considered in this study as they are sensitive toxicity markers that regulatory bodies such as the EPA use to set reference values (8).

Each endpoint was modeled in EPA's BMDS, version 2.6.0.1 where a best-fitting model was identified, with its associated BMD and BMDL values. The BMD is defined as a dose or concentration that produces a predetermined change in the response rate of an adverse effect, and the BMDL is the lower confidence limit of the BMD (9). Best-fit BMDs and BMDLs for every endpoint in each NTP study were identified, and then the minimum (and median where applicable) BMD and BMDL were determined for each chemical and study duration (chronic and short-term).

Through orthogonal regression techniques, the relationship between the observed minimum (and median) non-cancer chronic and short-term BMD10s and BMDL10s (and BMDL50 where applicable) in the chemicals sampled were assessed. BMD10 is defined as the dose associated with 10% response adjusted for background, BMD50 is associated with a 50% adjustment (9), and BMDLs refer to the corresponding lower limit of a one-sided 95% confidence interval on the BMD. BMDs and BMDLs were used for this analysis instead of reference doses to which uncertainty factors (UFs) are generally applied. This was done to minimize the variability and subjectivity that can result from application of UFs, which has the potential to result in inconsistent human health determinations. Further, BMDs and BMDLs are more consistent than other HHRVs (*e.g.*, NOAELs) because there are not as dependent on study design or power.

The findings indicate that short-term rodent studies can reasonably provide a quantitative estimate of chronic non-cancer BMDs and BMDLs. This may allow for faster development of points of departure (PODs), *i.e.* the estimated dose near the lower end of the observed range (9), for chemicals of concern in the environment that lack chronic toxicity data.

## 2 Methods

The following section provides the methodology used to develop modeling techniques to predict chronic toxicity levels from short-term toxicity data. Chronic and short-term test results reported by the NTP between 2000 and 2012 were used. The dose-response results for all non-neoplastic endpoints as well as organ and body weight, identified as significant by study pathologists, were modeled using the EPA's BMDS, Version 2.6.1. Best-fit BMD10, BMDL10, and where applicable BMD50 and BMDL50s for each significant non-neoplastic lesion, body weight (BW) and organ weight (OW) were identified. Then the minimum and median best-fit BMDs and BMDLs for each chemical were identified. This section will describe this process in detail.

### 2.1 Data collection

Data collection began by extracting publicly available TRs associated with NTP chronic Toxicity and Carcinogenicity Studies from their website (18). Most TRs include summaries of major findings associated with short-term studies for the same chemical substance. In addition, toxicity reports can be accessed which describe short-term studies in more detail (19).

All NTP TRs, and where necessary toxicity reports, were reviewed for specified chemical inclusion criteria; 41 chemicals met the criteria. For a list of chemicals that met the criteria refer to Table 3 in the Results section. The inclusion criteria were as follows: (1) the availability of both short-term and chronic toxicological studies; (2) exposure routes which include feed, drinking water, and/or gavage; (3) TR final publication date between 2000 and 2012. These dates were selected to ensure data both sufficient number of studies and that TRs were published as final publications.

For TRs that met these criteria, the next step was to identify endpoints reported as significant findings in the abstract of each TR. Only those endpoints are used in this study because they are assumed to be the significant health effects based on consistent criteria applied at NTP. This study assessed non-neoplastic lesions, BWs and OWs. For dichotomous endpoints (non-neoplastic lesions), a dataset within the context of this study includes the count of subjects (including species, sex and endpoint) within each exposure group where the pathologist identified a particular non-neoplastic lesion. For continuous endpoints a dataset is defined as the mean, standard deviation, and number of animals for BW or OW for each dose group.

Doses were generally reported by NTP in mg/kg of BW. Therefore, each dataset consisted of the following:

a. Chemical and TR

b. Duration (chronic and short-term)

    i. Chronic is approximately 2 years

    ii. Short-term is approximately 13 weeks

c. Species (F344/N and Wistar rat; B6C3F$_1$ mice)

d. Sex (male and female)

e. Dose group (*e.g.,* 0, 5, 10, 25, 55, 90 mg/kg per BW)

f. Non-cancer health effect:

  i. Continuous endpoints: average final BW or OW at the various dose group, total number of animal test subjects at each dose group, and the standard error of the mean final weight at each dose group.

  ii. Dichotomous endpoints: count of animals with non-neoplastic lesions out of total at each dose group

For continuous data points the adverse dose-response direction must be recorded as this is an important modeling assumption for continuous endpoints. Generally BW was observed to go down as dose increased but there were a few instances where BW actually increased following an exposure. Once the data collection was complete, each dataset was then fit with between five and eight possible models to determine the best-fit model.

## 2.2 Batch-processing and modeling in BMDS

The next step was to model all datasets and to identify a best-fit model. While BMD10 and BMDL10 were identified for all dataset types (*i.e.* continuous and dichotomous), BMD50 was only modeled for dichotomous datasets. This is because modeling BMD50 and BMDL50 would be associated with an average OW or BW increase or decrease of 50%. This would inevitably lead to estimates that are not physically possible for an animal to sustain.

The datasets were modeled in BMDS Version 2.6.1. Due to the large number of datasets to be modeled, batch-processing techniques were developed to capture, convert and create session files for the large amount of data. These methods were created using R code and Visual Basic programming to splice the datasets and save them in comma separated values format, then to create large text files to import datasets and create one large session file for BMDS to read. The session files were programmed to fit each dataset with various dose-response models.

Based on the BMDS user guidance the dose-response models fit to dichotomous datasets were: 1) Gamma, 2) Logistic, 3) Log-Logistic, 4) Log-Probit, 5) Multistage in multiple orders, 6) Probit, 7) Weibull, and 8) Quantal-Linear. For continuous data the dose-response curves were: 1) Exponential, 2) Hill, 3) Linear, 4) Polynomial in multiple orders, and 5) Power. (20) For continuous data, both modeled and constant variability were tested for every dataset within each chemical of interest (18). Then a series of post-modeling logical arguments based on the BMDS user-manual were applied to the results to identify whether modeled or constant variability was appropriate. The logical arguments applied are based on BMDS user guidance for Tests 2 and 3, which states:

> *Test 2: Tests the null hypothesis that variances are homogeneous. If this test fails to reject the null hypothesis, the simpler constant variance model may be appropriate.* [Otherwise, non-constant variance option was selected in the session title to remodel the dose-response relationship]
> *Test 3: Tests the null hypothesis that the variances are adequately modeled. If this test fails to reject the null hypothesis, it may be inferred that the variances have been modeled appropriately.*
> (20)

The outputs of the BMDS modeling process are Excel spreadsheets which were programmed to report the following information: dataset (i.e. health endpoint), model type, benchmark response (e.g. 0.1 or 0.5), measurement of increased risk (i.e. “extra risk”), BMD, BMDL, p-values for goodness-of-fit, Akaike Information Criterion (AIC), scaled residuals for dosed and control groups (21). These raw results were next assessed to identify the best-fit model for each dataset.

### 2.3 Filters

Once all datasets were modeled in BMDS, a best-fit model and thus best-fit

BMDs and BMDLs could be identified for each endpoint. Since BMDS is typically used on a smaller scale, guidance is not provided in the BMDS user manual for mass processing the selection of a best-fit model when multiple datasets need to be assessed simultaneously. However, by turning the BMDS user guidance and other research findings into a series of logical arguments, or *filters*, and applying them to all datasets simultaneously, best-fit results could be identified efficiently.

It is important to note that any data removed as a result of these filters was tracked, to identify endpoints for which no model could be fitted. The following provides the series of operationalized filters applied to the spreadsheets of compiled results.

Filter 1. If any modeling error was reported, the data was removed and tracked separately. This includes actual error indications, as well as physically impossible findings (*e.g.* BMD= -999999). Physically impossible findings are assumed to be the result of data that simply cannot be modeled by BMDS.

Filter 2. Examine the goodness-of-fit p-value and chi-squared residuals for each model generated. The p-value provides an indication of how well the model fits the observed dose-response data. The scaled residuals provide an indication of how well a model fits the dose group closest to the calculated BMD or BMDL. A model was retained if the p-value was greater than 0.1 *and* the absolute value of the chi-squared for residuals for each doses group were all less than or equal to 2.0. Any model not meeting these criteria is removed and tracked separately. (20)

Filter 3. If more than one model was found to meet the best-fit criteria, the Akaike Information Criterion (AIC) was assessed within the remaining model choices for each dataset. According to the BMDS user's guide, the AIC can be used to compare different types of models that use a similar fitting method, with the lowest AIC presumed to be the better fitting model. The AIC should only be used in cases of models with BMDL values within a threefold difference. (20)

Filter 4. After finding the minimum AIC, some datasets still had multiple best-fit models. According to the BMDS user guide, at this point visual assessment and a review of literature for assessing realistic results should be implemented to determine the best fit amongst the remaining models (20). However, with the vast quantity of data being examined this approach was not practical. Ringblom, *et al.* found that BMD to BMDL ratios usually fell between 3 and 10 (22). So datasets having BMD to BMDL ratios above 10 are not typical of a biological response (22), and any such models were removed.

Filter 5. The remaining best-fit models were then averaged together, although generally these instances resulted in BMD and BMDLs which were virtually or completely identical. (22) (23)

Filter 6. After the application of the first five filters, each *dataset* (i.e. endpoint) had a single (or averaged) best-fit model with a predicted BMD and BMDL (with the exception of a few for which no model could be fitted; these were also tracked). With numerous endpoints identified for each chemical and duration, the next step was to determine the minimum (or median), best-fit BMD and BMDL for each chemical within a chemical and study duration (*i.e*. chronic and short-term).

Although BMDLs are commonly used by regulatory bodies such as the EPA to set RfDs (24), BMD was also analyzed because it has the potential to be less influenced by study design and power. Similarly, BMD10 and BMDL10 versus

BMD50 and BMDL50, and medians versus minimums were considered. It was hypothesized that BMD50 and BMDL50 and/or median values could be a more robust POD. Continuous and dichotomous data were analyzed separately and later combined (referred to as *absolute*) to determine if dataset type influenced the final determinations. However, median BMD50s and BMDL50s and median absolute values were not determined to ease computation since these were not a primary research objective of the study. BMD50 and BMDL50 was not determined for continuous data because a 50% increase in risk for BW or OW would not provide a biologically realistic response. Therefore the following PODs were analyzed:

- Minimum dichotomous: BMD10, BMDL10, BMD50 and BMDL50
- Minimum continuous: BMD10, BMDL10
- Minimum absolute: BMD10 and BMDL10
- Median dichotomous: BMD10, BMDL10, BMD50 and BMDL50
- Median continuous: BMD10 and BMDL10

The objective is to relate short-term and chronic BMDs and BMDLs for each chemical. However, it was possible that all endpoints could be filtered out for certain chemicals in one or the other duration. This was the case for three chemicals, and the results are represented in Table 3

### 2.4 Relation between chronic and short-term BMDs and BMDLs

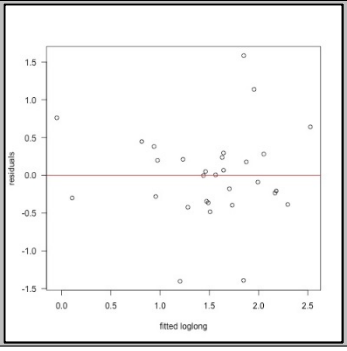
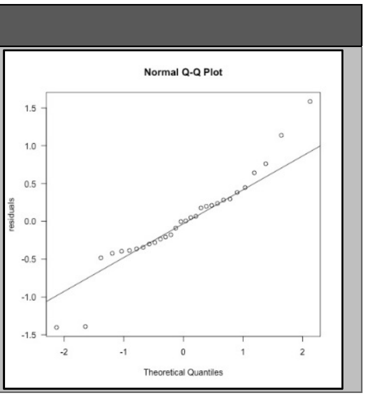
Next, the relations between chronic and short-term BMDs and BMDLs were examined for each of the following:

- Minimum dichotomous: BMD10, BMDL10, BMD50 and BMDL50
- Minimum continuous: BMD10, BMDL10

- Minimum absolute: BMD10 and BMDL10
- Median dichotomous: BMD10, BMDL10, BMD50 and BMDL50
- Median continuous: BMD10 and BMDL10

Scatter plots were made for all of the data listed above in log-log and untransformed scales (refer to Figure 2 and Figure 3 in the Results section). The relation of chronic to short-term data was quantified using orthogonal regression (see below). Residuals from these regressions were examined using residual plots and normal quantile plots augmented by the Shapiro-Wilks test (Table 1 and Table 2). Both untransformed and log-transformed values were examined. The log-transformation was unequivocally better. Chronic and short-term log-transformed BMDs and BMDLs are linearly related with approximate normality and homogeneity of variances (Table 1 and Table 2) (25). As indicated in the sample of findings below, after log-transformation of BMD and BMDL, p-values for the Shapiro-Wilk test were consistently above 0.05 and the Q-Q plots were roughly linear. Therefore, it was determined that a log-linear methodology was the most appropriate.

*Table 1: Sample log diagnostic results: minimum, best-fit BMDLs.*

| Minimums | | RESIDUAL PLOTS | |
|---|---|---|---|
| BMDL10c | W = 0.940,<br>p-value = 0.0894 |  |  |

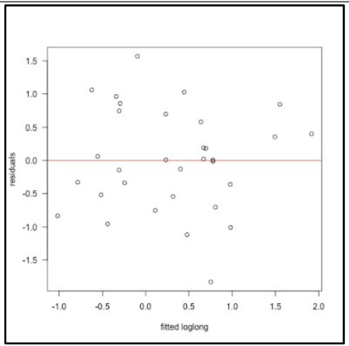
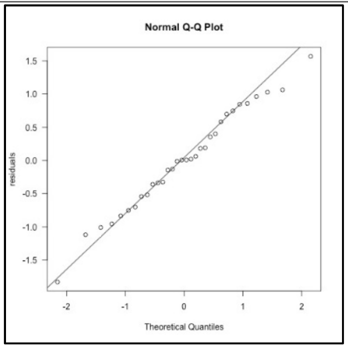
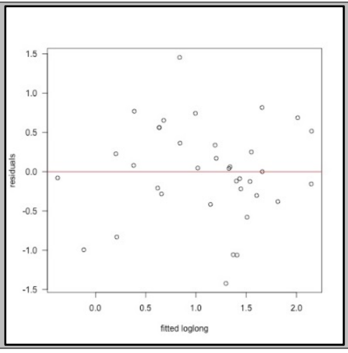
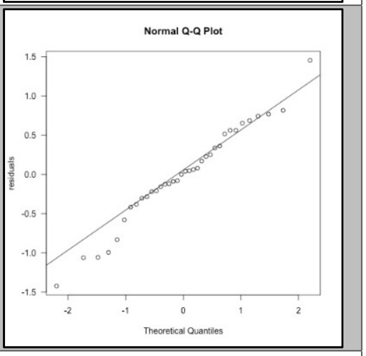
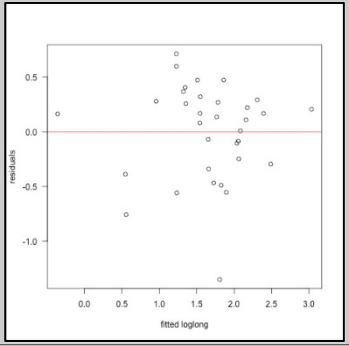
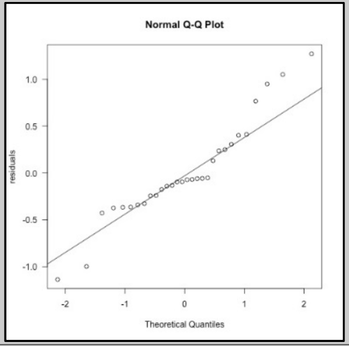
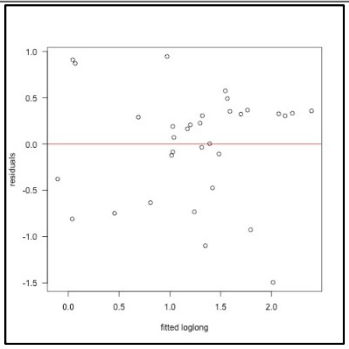
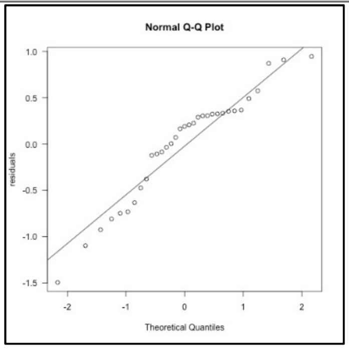
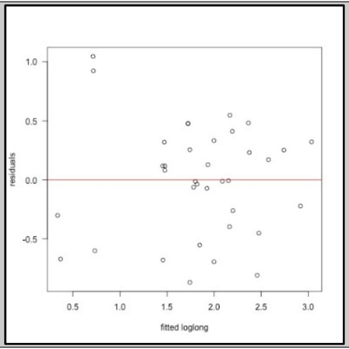
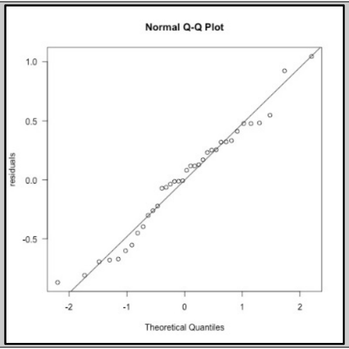
|  |  |  |  |
|---|---|---|---|
| BMDL10d | W = 0.988,<br>p-value = 0.9674 |  |  |
| BMDL50d | W = 0.979,<br>p-value = 0.714 |  |  |
| “c”= continuous; “d”= dichotomous; “a”= absolute |  |  |  |

*Table 2: Sample log diagnostic results: median, best-fit BMDLs.*

| MEDIANS |  | RESIDUAL PLOTS |  |
|---|---|---|---|
| BMDL10c | W = 0.9314,<br>p-value = 0.05349 |  |  |
| BMDL10d | W = 0.93618,<br>p-value = 0.05265 |  |  |

|  |  |  |  |
|---|---|---|---|
| BMDL50d | W = 0.97197,<br>p-value = 0.4817 |  |  |
| “c”= continuous; “d”= dichotomous; “a”= absolute |  |  |  |

Orthogonal regression was used because the predictor (short-term BMD or BMDL) is a random variable, i.e., there is error in both variables. Accurate estimation requires knowledge of the ratio between the variances of ‘X’ and 'Y', which cannot be estimated consistently from these data alone. For these data it was assumed the ratio of variances (involving measurement and equation error) is 1. This is a plausible approximation based upon the observations (Figure 2) and the similarity of the processes generating the 'X' and 'Y' data. Least-squares regressions were also estimated; the slopes were biased downward, leading to over-prediction at low values and under-prediction at high values. Accurate confidence intervals cannot be provided without knowledge of the ratio of variances or replicate observations (26). However, the dispersion of observations in Figure 2 suggests that prediction of a single new observation could entail a prediction error of between 10-fold and 100-fold. This is also reflected in the least-squares regressions, of which one example is shown in Figure 4 (note that this example was selected merely because it is representative of the results). (27) (28) (29) (30) (31)

Orthogonal regressions were done in R code (32) and the estimates are reported in Table 9 of the Results section. The resulting regression lines are included in Figure 2 and Figure 3.

## 3 Results

The goal of this study was to assess the ability of short-term toxicity results to reasonably predict chronic non-cancer health effect levels from short-term toxicity data. This section provides the results of the investigation including: the chemicals from the NTP TRs which met the criteria of the chemical inclusion assessment; the minimum, best-fit BMDLs per chemical-duration; and the scatter plots and correlations resulting from the statistical assessments of the observed distributions.

### 3.1 Chemical inclusion results

Table 3 describes those chemicals that met the chemical inclusion criteria: (1) availability of both short-term and chronic toxicological studies; (2) exposure routes which include feed, drinking water, and/or gavage; (3) TR final publication date between 2000 and 2012. While all of the chemicals listed in Table 3 met the chemical inclusion criteria, a particular duration was absent for three out of the 41 chemicals after the filtering described in the Methods section. This can be attributed to: (1) no significant non-cancer health effects were reported in one study duration of a particular chemical; (2) no best-fit model was found; (3) best-fit models were identified but not indicative of a biological response, for example the BMD to BMDL ratio was greater than 10.

*Table 3: NTP analyzed chemicals meeting chemical inclusion criteria.*

| Report # | Toxicity report* | Chemical | CAS | Exposure |
|---|---|---|---|---|
| TR470 | | Pyridine | 110-86-1 | Drinking Water |
| TR476 | | Primidone | 125-33-7 | Feed |
| TR491 | | Methyleugenol | 93-15-2 | Gavage |
| TR493 | | Emodin | 518-82-1 | Feed |
| TR494 | | Anthraquinone | 84-65-1 | Feed |
| TR495 | | Sodium Nitrite | 7632-00-0 | Drinking Water |
| TR497 | TOX-47 | Methacrylonitrile | 126-98-7 | Gavage |
| TR498 | TOX-23 | p-Nitrotoluene | 99-99-0 | Feed |
| TR501 | | p,p'-Dichlorodiphenyl Sulfone | 80-07-9 | Feed |

|  |  |  |  |  |
|---|---|---|---|---|
| TR504 | TOX-44 | o-Nitrotoluene | 88-72-2 | Feed |
| TR505 |  | Citral | 5392-40-5 | Feed |
| TR506 |  | Acrylonitrile | 107-13-1 | Gavage |
| TR508 | Tox-27 | Ridilline | 23246-96-0 | Gavage |
| TR509 |  | 2,4-Hexadienal | 142-83-6 | Gavage |
| TR510a |  | Urethane + 0%Ethanol | 51-79-6 | Drinking Water |
| TR510b* | TOX-52 | Urethane + 5%Ethanol | 51-79-6 & 64-17-5 | Drinking Water |
| TR511 |  | Dipropylene Glycol | 25265-71-8 | Drinking Water |
| TR512* |  | Elmiron® | 37319-17-8 | Gavage |
| TR514 |  | trans-Cinnamaldehyde (Microencapsulated) | 14371-10-9 | Feed |
| TR516 |  | 2-Methylimidazole | 693-98-1 | Feed |
| TR533 |  | Benzophenone | 119-61-9 | Feed |
| TR535 | TOX-67 | 4-Methylimidazole | 822-36-6 | Feed |
| TR537 |  | Dibromoacetic Acid | 631-64-1 | Drinking Water |
| TR540 |  | Methylene Blue Trihydrate | 7220-79-3 | Gavage |
| TR541 |  | Formamide | 75-12-7 | Gavage |
| TR544 |  | Dibromoacetonitrile | 3252-43-5 | Drinking Water |
| TR546 | TOX-72 | Sodium Dichromate Dihydrate | 7789-12-0 | Drinking Water |
| TR549 |  | Bromochloroacetic Acid | 5589-96-8 | Drinking Water |
| TR550 | TOX-9 | Cresols | 1319-77-3 | Feed |
| TR551 |  | Isoeugenol | 97-54-1 | Gavage |
| TR554* |  | 5-(Hydroxymethyl)-2-Furfural | 67-47-0 | Gavage |
| TR557 |  | B-Myrcene | 123-35-3 | Gavage |
| TR558 | TOX-65 | 3,3',4,4'-Tetrachloroazobenzene | 14047-09-7 | Gavage |
| TR560 |  | Androstenedione | 63-05-8 | Gavage |
| TR562 |  | Goldenseal Root Powder | GOLDENSEALRT | Feed |
| TR563 |  | Pulegone | 89-82-7 | Gavage |
| TR565* |  | Milk Thistle Extract | 84604-20-6 | Feed |
| TR570 |  | Alpha,Beta-Thujone | 76231-76-0 | Gavage |
| TR571 |  | Kava Kava Extract | 9000-38-8 | Gavage |
| TR575 |  | Acrylamide | 79-06-1 | Drinking Water |
| TR579 |  | N,N-Dimethyl-P-Toluidine | 99-97-8 | Gavage |
| * Not represented in final results. |  |  |  |  |

3.2 Best-fit, minimum and median BMD and BMDLs

Best-fit BMDs and BMDLs per endpoint were determined using the filters described in the Methods section. Next, the minimum (or median) best-fit BMD or BMDL within each chemical-duration category was identified. These results are reported below in Table 4 through Table 8.

Table 4 represents the *absolute* minimum, best-fit BMDL per chemical-duration. Minimum, best-fit BMDLs were also separately determined per chemical-duration for continuous and dichotomous endpoints only.  These results can be viewed in the scatter plots reported in Figure 2 and Figure 3, but  not represented in tabular format for conciseness.

*Table 4: Best-fit, minimum BMD10 and BMDL10 absolute results.*

| NTP TR # | Best-Fit, Minimum BMD10 Absolute | | | | Best-Fit, Minimum BMDL10 Absolute | | | |
|---|---|---|---|---|---|---|---|---|
| | Chronic (mg/kg BW) | Endpoint type* | Short (mg/kg BW) | Endpoint type* | Chronic (mg/kg BW) | Endpoint type* | Short (mg/kg BW) | Endpoint type* |
| TR470 | 18.69 | c | 6.32 | c | 14.88 | c | 4.03 | c |
| TR476 | 2.15 | d | 4.18 | d | 1.74 | d | 2.94 | d |
| TR491 | 4.84 | c | 0.64 | d | 2.86 | c | 0.48 | d |
| TR493 | 457.51 | d | 235.32 | c | 207.62 | d | 176.44 | c |
| TR494 | 4.15 | d | 0.70 | d | 2.76 | d | 0.46 | d |
| TR495 | 1.79 | d | 26.21 | d | 1.27 | d | 21.01 | d |
| TR497 | 39.82 | c | 7.46 | d | 27.29 | c | 6.09 | d |
| TR498 | 0.06 | d | 0.51 | d | 0.04 | d | 0.29 | d |
| TR501 | 8.45 | c | 4.88 | c | 4.73 | c | 3.75 | c |
| TR504 | 0.39 | d | 2.98 | d | 0.23 | d | 1.94 | d |
| TR505 | 111.73 | c | 585.91 | c | 93.62 | c | 419.14 | c |
| TR506 | 9.12 | d | 34.84 | d | 5.85 | d | 19.11 | d |
| TR508 | 0.14 | d | 0.16 | d | 0.08 | d | 0.09 | d |
| TR509 | 5.01 | d | 63.84 | d | 4.15 | d | 37.87 | d |
| TR510a | 40.11 | d | 8.86 | d | 29.58 | d | 0.95 | d |
| TR511 | 14.98 | d | 24.28 | d | 7.55 | d | 14.44 | d |
| TR514 | 150.24 | c | 385.94 | d | 81.01 | c | 271.26 | d |
| TR516 | 1.49 | d | 66.05 | d | 0.94 | d | 38.46 | d |
| TR533 | 1.24 | d | 32.58 | d | 0.23 | d | 6.83 | d |
| TR535 | 8.56 | d | 25.62 | d | 4.89 | d | 13.10 | d |
| TR537 | 15.39 | d | 5.08 | d | 8.52 | d | 2.93 | d |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| TR540 | 3.64 | d | 0.50 | d | 2.76 | d | 0.16 | d |
| TR541 | 22.47 | d | 24.55 | d | 16.64 | d | 11.82 | d |
| TR544 | 10.79 | c | 13.48 | c | 7.22 | c | 13.19 | c |
| TR546 | 0.48 | d | 0.70 | d | 0.35 | d | 0.46 | d |
| TR549 | 11.34 | d | 21.58 | d | 7.22 | d | 13.10 | d |
| TR550 | 3.01 | d | 20.93 | d | 1.87 | d | 5.26 | d |
| TR551 | 7.46 | d | 41.81 | d | 6.07 | d | 18.99 | d |
| TR557 | 0.03 | d | 0.06 | d | 0.01 | d | 0.04 | d |
| TR558 | 0.48 | d | 0.28 | c | 0.32 | d | 0.08 | c |
| TR560 | 0.37 | d | 1.95 | d | 0.26 | d | 0.57 | d |
| TR562 | 87.45 | d | 211.30 | c | 70.22 | d | 117.75 | c |
| TR563 | 0.95 | d | 6.86 | d | 0.59 | d | 3.91 | d |
| TR570 | 10.45 | d | 1.01 | d | 4.20 | d | 0.41 | d |
| TR571 | 0.14 | d | 0.19 | c | 0.09 | d | 0.14 | c |
| TR575 | 0.30 | d | 31.83 | d | 0.08 | d | 17.44 | d |
| TR579 | 0.92 | d | 1.25 | d | 0.35 | d | 0.39 | d |

**c= continuous; d=dichotomous*

Table 5 represents the best-fit, minimum BMD50 results. These results were only analyzed for dichotomous endpoints because the BMD50 for a continuous endpoint does not represent a biologically likely finding. For example, the BMD or BMDL at which an animal would lose 50% of its BW would likely result in the animals death.

*Table 5: Best-fit, minimum BMD50 and BMDL50 results.*

| | Best-Fit, Minimum BMD50* (mg/kg BW) | | Best-Fit, Minimum BMDL50* (mg/kg BW) | |
|---|---|---|---|---|
| NTP TR# | Chronic | Short-term | Chronic | Short-term |
| TR470 | 14.34 | 27.20 | 11.58 | 11.39 |
| TR476 | 40.44 | 28.68 | 33.65 | 20.16 |
| TR491 | 29.95 | 28.71 | 23.54 | 20.97 |
| TR493 | 13.10 | 45.19 | 2.23 | 41.72 |
| TR494 | 7.60 | 7.33 | 2.69 | 0.76 |
| TR495 | 120.54 | 602.96 | 97.83 | 483.29 |
| TR497 | 26.35 | 50.96 | 19.11 | 40.69 |
| TR498 | 20.83 | 22.03 | 15.69 | 3.19 |

| | | | | |
|---|---|---|---|---|
| TR501 | 20.83 | 22.03 | 15.69 | 3.19 |
| TR504 | 30.18 | 62.44 | 22.10 | 45.34 |
| TR506 | 60.00 | 38.71 | 23.35 | 31.84 |
| TR508 | 0.36 | 0.18 | 0.35 | 0.11 |
| TR509 | 28.91 | 85.20 | 25.74 | 63.72 |
| TR510a | 263.88 | 12.21 | 194.60 | 6.25 |
| TR510b | 352.22 | 177.90 | 296.58 | 95.27 |
| TR511 | 91.06 | 162.06 | 45.57 | 95.83 |
| TR512 | 22.50 | 11.10 | 15.98 | 6.34 |
| TR514 | 548.79 | 423.06 | 497.09 | 309.29 |
| TR516 | 13.39 | 74.62 | 8.46 | 57.84 |
| TR533 | 7.00 | 190.07 | 2.06 | 36.72 |
| TR535 | 29.44 | 85.77 | 19.90 | 79.42 |
| TR537 | 138.53 | 18.30 | 54.55 | 10.55 |
| TR540 | 19.07 | 4.50 | 14.26 | 1.40 |
| TR541 | 68.75 | 92.67 | 63.36 | 66.93 |
| TR546 | 3.29 | 4.62 | 2.57 | 3.03 |
| TR549 | 29.71 | 39.54 | 25.12 | 33.04 |
| TR550 | 27.12 | 106.41 | 16.86 | 47.31 |
| TR551 | 49.11 | 212.40 | 27.13 | 159.83 |
| TR557 | 0.10 | 0.40 | 0.08 | 0.26 |
| TR558 | 4.29 | 2.31 | 2.88 | 1.37 |
| TR560 | 3.33 | 5.00 | 2.37 | 3.43 |
| TR562 | 575.34 | 524.10 | 461.99 | 487.90 |
| TR563 | 8.58 | 22.99 | 5.34 | 17.31 |
| TR570 | 26.80 | 9.09 | 21.45 | 3.70 |
| TR571 | 0.27 | 1.00 | 0.24 | 0.79 |
| TR575 | 2.74 | 35.96 | 0.75 | 28.81 |
| *Only dichotomous endpoints were analyzed for BMD50 and BMDL50 results. | | | | |

Table 6 represents the best-fit median results for BMD10 and BMDL10, dichotomous datasets only; and Table 7 represents the results for the best-fit BMD10 and BMDL10, continuous datasets only. No absolute determination was made for the median data because medians were not the primary focus of the study. Since BMDS modeling format dictates the segregation of the dataset types and median findings were not primary to the research question the results remain in that format.

*Table 6: Best-fit, median BMD10 and BMDL10 dichotomous datasets.*

| NTP TR# | Best-fit, Median BMD10 (mg/kg BW)<br>Dichotomous datasets | | Best-fit, Median BMDL10 (mg/kg BW)<br>Dichotomous datasets | |
|---|---|---|---|---|
| | Chronic | Short-term | Chronic | Short-term |
| TR476 | 28.12 | 55.91 | 21.86 | 34.21 |
| TR491 | 60.75 | 109.28 | 42.42 | 58.99 |
| TR493 | 557.08 | 2635.12 | 345.22 | 1563.73 |
| TR494 | 152.14 | 26.40 | 83.06 | 16.26 |
| TR495 | 5.11 | 1435.69 | 3.33 | 776.66 |
| TR497 | 213.84 | 225.70 | 132.17 | 135.48 |
| TR498 | 13.60 | 8.54 | 9.56 | 5.71 |
| TR504 | 4.83 | 59.27 | 3.24 | 44.02 |
| TR506 | 14.55 | 34.86 | 7.85 | 19.13 |
| TR508 | 0.70 | 4.75 | 0.51 | 2.39 |
| TR509 | 23.08 | 103.13 | 19.08 | 57.68 |
| TR510a | 161.56 | 482.89 | 105.64 | 240.46 |
| TR511 | 747.04 | 6731.90 | 567.86 | 3159.71 |
| TR514 | 292.88 | 1838.23 | 250.30 | 951.77 |
| TR516 | 115.55 | 224.64 | 88.32 | 160.87 |
| TR533 | 9.89 | 548.21 | 7.40 | 341.61 |
| TR535 | 45.74 | 158.67 | 23.81 | 107.57 |
| TR537 | 15.56 | 113.89 | 8.84 | 85.34 |
| TR540 | 14.84 | 37.23 | 8.79 | 20.04 |
| TR541 | 33.69 | 139.75 | 24.98 | 76.91 |
| TR546 | 2.06 | 12.29 | 1.50 | 8.85 |
| TR549 | 19.77 | 28.29 | 12.94 | 20.88 |
| TR550 | 203.74 | 508.78 | 135.53 | 305.70 |
| TR551 | 225.70 | 197.80 | 114.27 | 146.68 |
| TR557 | 0.21 | 0.93 | 0.17 | 0.51 |
| TR558 | 14.16 | 1.68 | 8.98 | 0.52 |
| TR560 | 12.80 | 1.95 | 8.71 | 0.57 |
| TR562 | 348.51 | 1911.32 | 274.76 | 1199.23 |

| | | | | |
|---|---|---|---|---|
| TR563 | 50.51 | 99.24 | 33.35 | 54.01 |
| TR570 | 36.41 | 61.20 | 25.65 | 37.90 |
| TR571 | 0.48 | 0.43 | 0.33 | 0.30 |
| TR575 | 2.78 | 139.72 | 1.79 | 65.81 |
| TR579 | 25.01 | 55.32 | 16.61 | 20.01 |

*Table 7: Best-fit, median BMD10 and BMDL10, continuous datasets*

| NTP TR# | Best-fit, Median BMD10 (mg/kg BW) Continuous datasets | | Best-fit, Median BMDL10 (mg/kg BW) Continuous datasets | |
|---|---|---|---|---|
| | Chronic | Short-term | Chronic | Short-term |
| TR470 | 18.69 | 12.93 | 14.88 | 4.466 |
| TR476 | 62.90 | 42.40 | 56.99 | 35.32 |
| TR491 | 20.10 | 350.98 | 12.447515 | 259.67 |
| TR493 | 1776.15 | 2322.70 | 1240.83 | 586.50 |
| TR494 | 494.67 | 49.69 | 313.92 | 27.38 |
| TR495 | 163.01 | 494.93 | 124.27 | 362.68 |
| TR497 | 39.82 | 48.75 | 27.29 | 32.67 |
| TR498 | 529.20 | 700.89 | 362.13 | 452.38 |
| TR501 | 15.486 | 36.34 | 11.01 | 24.03 |
| TR504 | 145.12 | 133.15 | 96.14 | 112.65 |
| TR505 | 176.90 | 601.01 | 139.75 | 494.53 |
| TR508 | 1.60 | 13.78 | 1.0346 | 9.84 |
| TR509 | 128.32 | 71.14 | 111.72 | 53.11 |
| TR510a | 168.38 | 293.92 | 80.18 | 219.45 |
| TR511 | 1918.72 | 3730.44 | 1649.11 | 3202.89 |
| TR514 | 386.47 | 1145.85 | 236.26 | 884.40 |
| TR516 | 189.485 | 668.91 | 138.07 | 564.34 |
| TR533 | 67.48 | 261.23 | 49.17 | 235.95 |
| TR535 | 83.49 | 157.23 | 67.16 | 123.70 |
| TR537 | 34.25 | 102.91 | 29.63 | 93.34 |
| TR540 | 28.16 | 26.28 | 18.20 | 20.03 |
| TR541 | 55.54 | 89.97 | 44.89 | 71.33 |
| TR544 | 14.50 | 13.48 | 9.89 | 13.19 |
| TR549 | 41.59 | 121.01 | 25.62 | 85.24 |
| TR558 | 10.03 | 0.33 | 5.55 | 0.088 |
| TR562 | 3130.91 | 531.57 | 2718.79 | 332.23 |

| | | | | |
|---|---|---|---|---|
| TR563 | 44.40 | 82.59 | 41.95 | 61.99 |
| TR570 | 17.17 | 47.78 | 13.35 | 29.34 |
| TR571 | 0.93 | 0.65 | 0.76 | 0.53 |
| TR579 | 38.01 | 97.30 | 27.78 | 63.64 |

Table 8 represents the best-fit, median BMD50 and BMDL50 results. As with the best-fit, minimum BMD50 and BMDL50 results, only dichotomous datasets were considered because the BMD50 or BMDL50 of continuous datasets would represent results that are not biologically likely to occur.

*Table 8: Best-fit, median BMD50 and BMDL50 results.*

| NTP TR# | Best-fit, Median BMD50 (mg/kg BW) Dichotomous datasets | | Best-fit, Median BMDL50 (mg/kg BW) Dichotomous datasets | |
|---|---|---|---|---|
| | Chronic | Short-term | Chronic | Short-term |
| TR470 | 42.28 | 44.72 | 36.16 | 35.39 |
| TR476 | 68.30 | 374.19 | 58.62 | 293.07 |
| TR491 | 140.29 | 689.90 | 87.16 | 326.59 |
| TR493 | 27.00 | 241.36 | 20.19 | 176.10 |
| TR494 | 866.60 | 359.47 | 521.82 | 296.40 |
| TR495 | 123.69 | 847.39 | 106.27 | 763.98 |
| TR497 | 27.13 | 169.82 | 19.73 | 110.53 |
| TR498 | 202.74 | 107.12 | 158.87 | 75.57 |
| TR501 | 202.74 | 107.12 | 158.87 | 75.57 |
| TR504 | 161.33 | 334.92 | 139.64 | 281.76 |
| TR506 | 109.42 | 39.46 | 37.38 | 33.04 |
| TR508 | 1.79 | 8.30 | 1.36 | 3.61 |
| TR509 | 70.45 | 119.81 | 60.52 | 101.09 |
| TR510a | 1109.75 | 623.04 | 704.79 | 542.47 |
| TR510b | 653.97 | 610.44 | 404.03 | 557.69 |
| TR511 | 2545.79 | 9128.47 | 2281.76 | 4259.63 |
| TR512 | 159.58 | 266.64 | 115.88 | 144.47 |
| TR514 | 548.79 | 3520.56 | 497.09 | 2965.76 |
| TR516 | 237.33 | 241.22 | 214.73 | 175.82 |
| TR533 | 53.05 | 752.43 | 44.54 | 719.75 |
| TR535 | 240.81 | 310.48 | 119.45 | 231.68 |

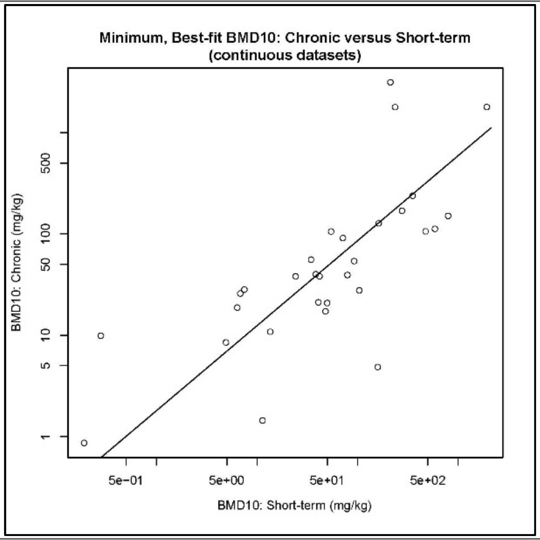
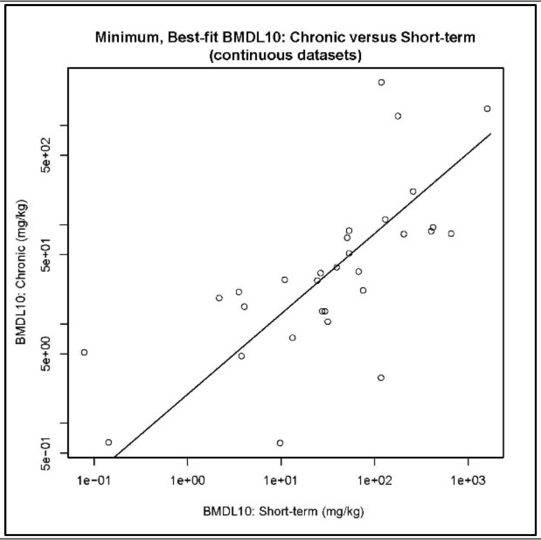
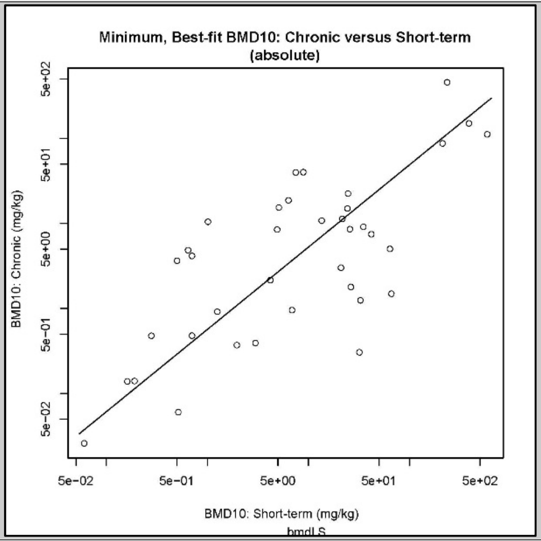
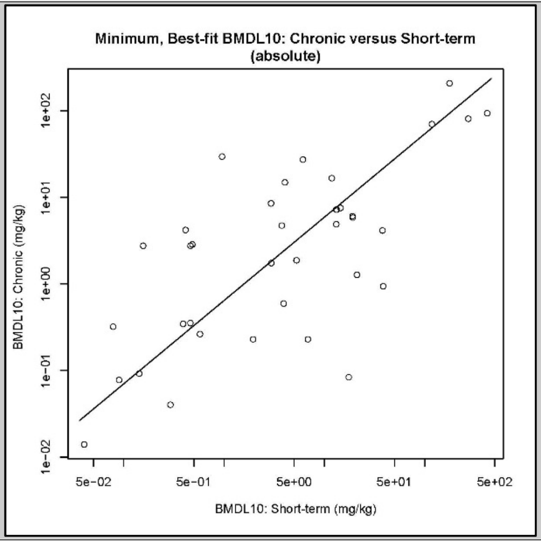
|  |  |  |  |  |
|---|---|---|---|---|
| TR537 | 140.05 | 122.29 | 61.48 | 95.90 |
| TR540 | 45.21 | 52.11 | 39.30 | 35.36 |
| TR541 | 88.37 | 170.40 | 71.26 | 139.78 |
| TR546 | 7.66 | 36.15 | 6.01 | 33.37 |
| TR549 | 132.74 | 39.99 | 61.82 | 34.77 |
| TR550 | 656.46 | 1271.83 | 561.08 | 1044.03 |
| TR551 | 740.88 | 441.09 | 402.60 | 319.67 |
| TR557 | 0.54 | 1.67 | 0.50 | 1.18 |
| TR558 | 92.41 | 6.59 | 57.63 | 3.40 |
| TR560 | 54.74 | 5.00 | 43.79 | 3.43 |
| TR562 | 1255.41 | 2159.47 | 989.74 | 1728.84 |
| TR563 | 113.98 | 127.44 | 99.50 | 79.92 |
| TR570 | 62.65 | 102.35 | 52.38 | 90.33 |
| TR571 | 1.34 | 1.56 | 1.09 | 1.07 |
| TR575 | 11.20 | 155.13 | 7.48 | 80.03 |

## 3.3 Results of orthogonal regressions

This section describes how the results presented in Table 4 through Table 8 were plotted including the trend lines for the orthogonal regression techniques. Chronic non-cancer toxicity data represents the dependent variable, and the short-term toxicity data represents the independent variable. Orthogonal regressions were applied to each scatter plot to examine the linear relationship because it is assumed that both variables have some measurement error (28). The resulting scatter plots and trend lines are reported in Figure 2 and Figure 3 and the correlations and statistical assessments for these plots can be found in Table 9. Since these data were determined to be log-log, it is represented by the form:

*y= 10^(Coefficient a+ Coefficient b*log(x)).*

*Figure 2: Minimums: best-fit, BMD and BMDL scatter plots (Log10)*

<table>
<tr><th>Best-fit BMD</th><th>Best Fit BMDL</th></tr>
<tr><td colspan="2">Minimum Results</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

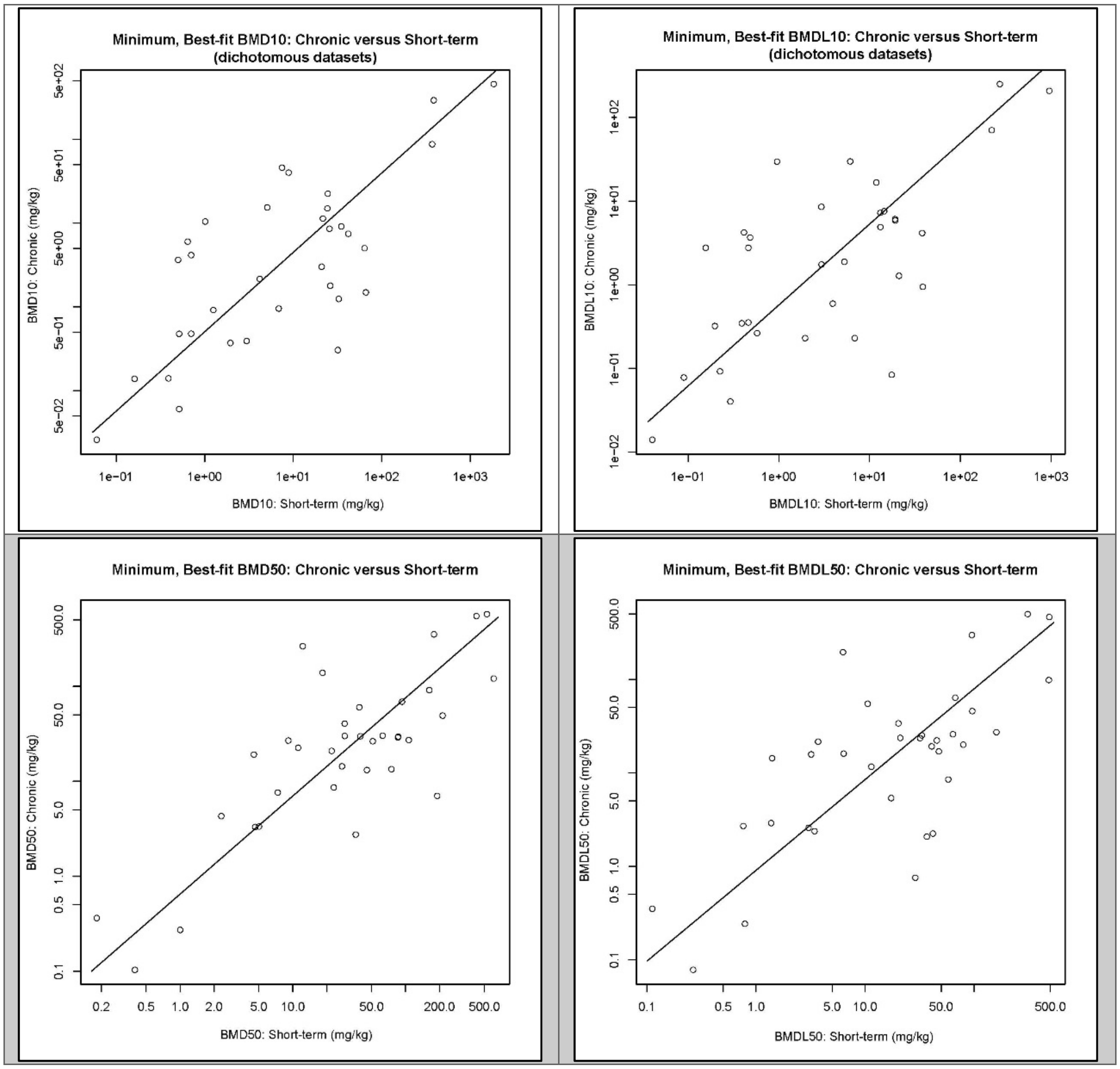
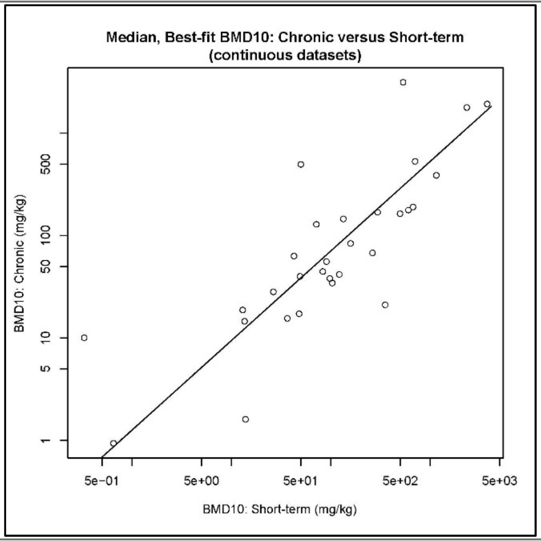
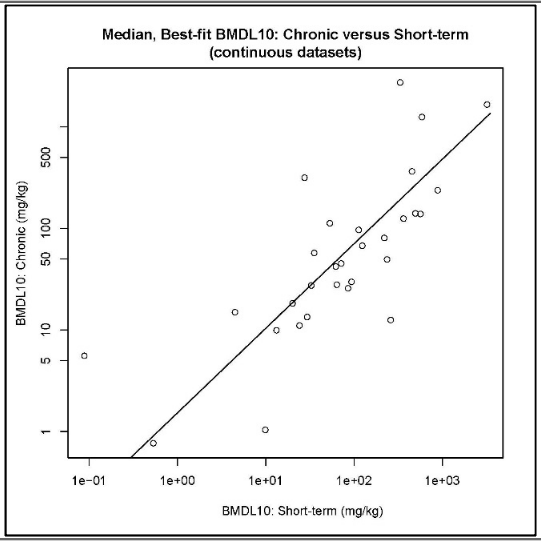
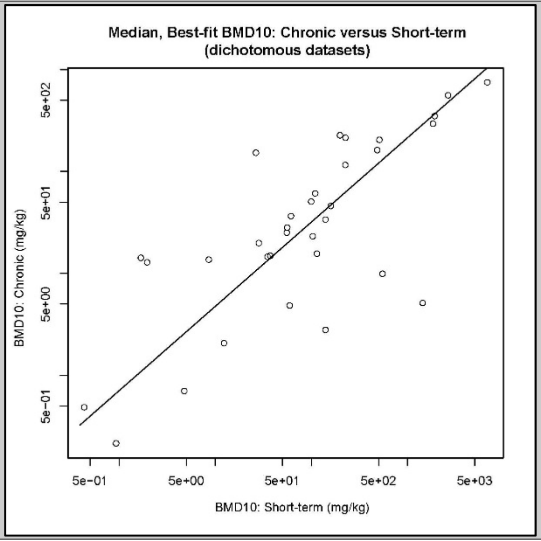
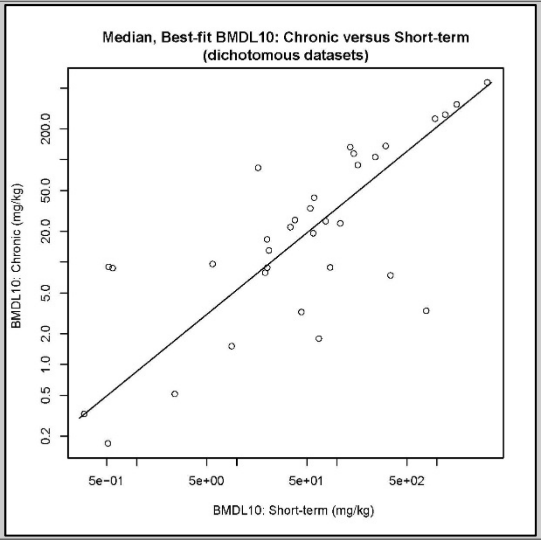

Minimum, Best-fit BMD10: Chronic versus Short-term
(dichotomous datasets)
BMD10: Chronic (mg/kg)
5e−02
5e−01
5e+00
5e+01
5e+02
1e−01
1e+00
1e+01
1e+02
1e+03
BMD10: Short-term (mg/kg)
Minimum, Best-fit BMDL10: Chronic versus Short-term
(dichotomous datasets)
BMDL10: Chronic (mg/kg)
1e−02
1e−01
1e+00
1e+01
1e+02
1e−01
1e+00
1e+01
1e+02
1e+03
BMDL10: Short-term (mg/kg)
Minimum, Best-fit BMD50: Chronic versus Short-term
BMD50: Chronic (mg/kg)
0.1
0.5
1.0
5.0
50.0
500.0
0.2
0.5
1.0
2.0
5.0
10.0
50.0
200.0
500.0
BMD50: Short-term (mg/kg)
Minimum, Best-fit BMDL50: Chronic versus Short-term
BMDL50: Chronic (mg/kg)
0.1
0.5
1.0
5.0
50.0
500.0
0.1
0.5
1.0
5.0
10.0
50.0
500.0
BMDL50: Short-term (mg/kg)

*Figure 3: Median results: Best-fit, BMD and BMDL scatter plot results (Log10)*

| Best-fit BMD | Best Fit BMDL |
|---|---|
| Median Results | |
|  |  |
|  |  |

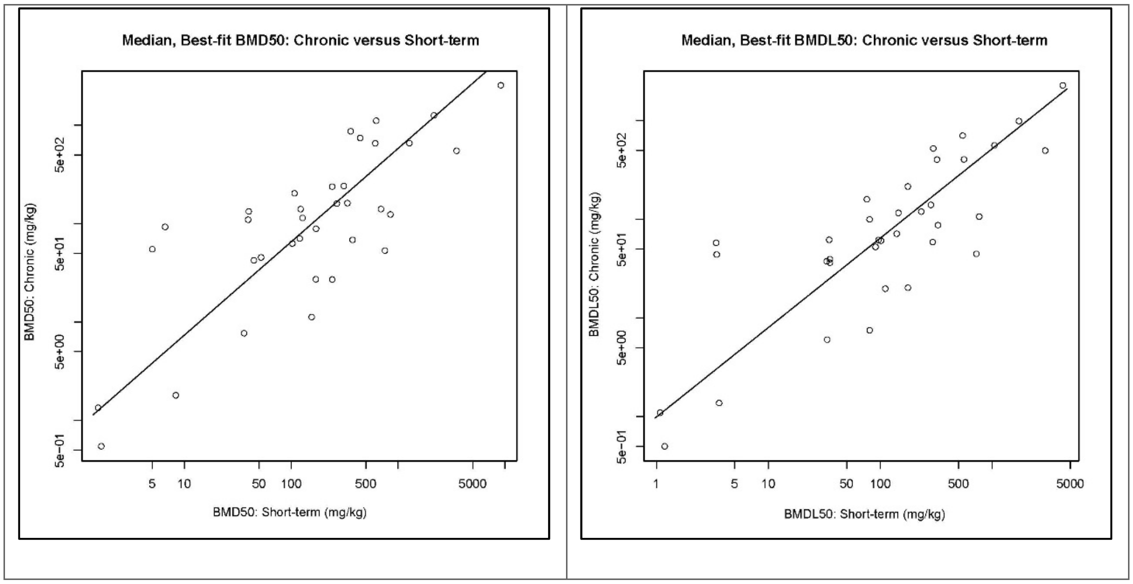


Figure 4 represents the least-square regressions for minimum BMDL10 chronic versus short-term (selected as an example because it is representative of results). The 90% confidence interval are represented with dotted lines, and the 90% prediction interval for a single prediction in dashed lines.  This figure highlights the amount of error (between 10-100 fold) associated with the predication of a single chemical from the data (i.e. the prediction interval).  Prediction intervals calculate where the next data point sampled is expected.  Confidence intervals assess how well the mean has been determined given a random sample.  Therefore, the prediction interval must account for both the uncertainty of the sample mean plus the uncertainty within the data scatter. So a prediction interval is always wider than a confidence interval, which in this study accounts for much of the error in predicting a single chemical's chronic BMD or BMDL. (33)

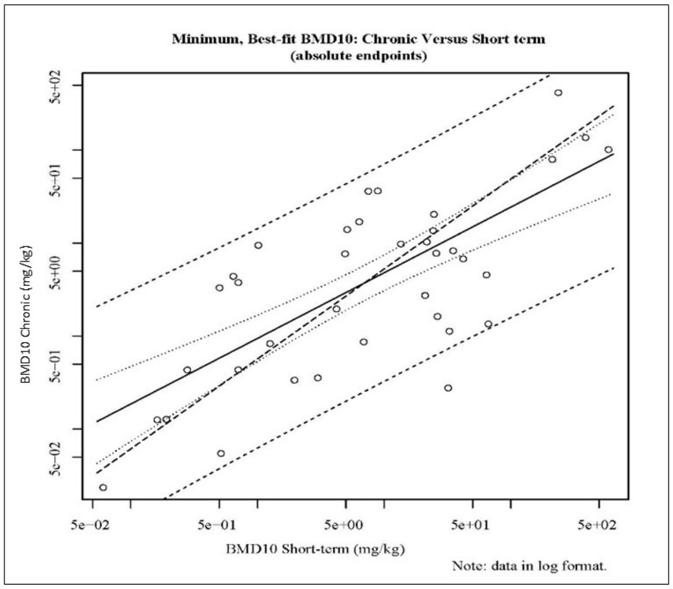


*Figure 4: Least-squares regression minimum BMDL10*

The trend lines reported in the scatter plots in Figures 2 and 3 correspond to function: *y= 10^(Coefficient a+ Coefficient b*log(x)),* where coefficient 'a' and 'b' and the confidence limits for the 'b' (*i.e.* slope) are reported for each parameter in Table 9. Results of the Pearson correlations ranged from 0.69 to 0.81; and Spearman correlations ranged from 0.59 to 0.84. All were significant at $p< 0.001$.

*Table 9: Orthogonal regression analysis statistical assessments*

| | | Minimum | | | Median | | |
|---|---|---|---|---|---|---|---|
| | | | Coefficients (UL, LL)* | | | Coefficients | |
| | POD | N | Intercept (a) | Slope (b) | N | Intercept (a) | Slope (b) |
| Absolute | BMD10 | 37 | -0.25 | 0.97 (0.69, 1.36) | | | |
| | BMDL10 | 37 | -0.19 | 0.96 (0.66, 1.39) | | | |
| Continuous | BMD10 | 30 | 0.26 | 0.84 (0.57, 1.21) | 30 | 0.10 | 0.87 (0.65. 1.16) |
| | BMDL10 | 30 | 0.29 | 0.81 (0.52, 1.22) | 30 | -0.15 | 0.83 (0.59, 1.16) |
| Dichotomous | BMD10 | 33 | -0.30 | 0.95 (0.66, 1.35) | 33 | -0.08 | 0.95 (0.58, 1.15) |
| | BMD50 | 36 | -0.19 | 1.03 (0.79, 1.36) | 36 | 0.18 | 0.83 (0.72. 1.25) |
| | BMDL10 | 33 | -0.24 | 0.97 (0.64, 1.45) | 33 | -0.07 | 0.80 (0.55, 1.11) |
| | BMDL50 | 36 | -0.04 | 0.97 (0.67, 1.39) | 36 | 0.00 | 0.91 (0.70, 1.17) |

**Upper Limit (UL); Lower Limit (LL)*

## 4 Discussion

This study investigated whether, in the absence of the necessary chronic non-cancer toxicity data, short-term non-cancer toxicity data can be used to predict chronic non-cancer health effect levels by focusing on the dose-response relationship instead of a critical effect. This is a novel approach. Based on a review of publically available chemical substitution and single chemical assessment frameworks there is typically a focus on hazard instead of dose or effect level.

In practice, a minimum BMDL10 is typically used to developed reference doses (RfDs) by such regulatory bodies as EPA (24). Selecting the health effect that occurred at the lowest dose (i.e. critical effect) is considered more health protective (24). However, recall that in this study, in addition to minimum BMD10 and BMDL10, other parameters were also analyzed that were hypothesized to be less influenced by study design and power which do not exist at the lower or upper boundary of the spectrum of data. Therefore, the following parameters were analyzed:

- Minimum dichotomous: BMD10, BMDL10, BMD50 and BMDL50
- Minimum continuous: BMD10, BMDL10
- Minimum absolute: BMD10 and BMDL10
- Median dichotomous: BMD10, BMDL10, BMD50 and BMDL50
- Median continuous: BMD10 and BMDL10

The findings indicate that regardless of the chemical or health endpoint, short-term toxicity studies reasonably provide a quantitative estimate of chronic PODs. This can allow for faster development of PODs for target chemicals in the environment that lack chronic toxicity data. The chronic and short-term BMDs and BMDLs were highly correlated in all cases, with Pearson correlation coefficients between 0.69 and 0.81 ($p < 0.001$). Further, the lower confidence limits of the orthogonal regression for b coefficients (*i.e.* slope) are all greater than zero, which is also an indication of good model fit. The scatter plots suggest log-log form and the Pearson and Spearman correlations strongly support that impression.

Based on the statistical findings, BMD and BMDL median and minimum best-fit non-cancer chronic and short-term data all appear to be strongly correlated. In practice minimum values are utilized to derive reference values (24), however this study investigated other descriptions of toxicity, hypothesizing that, for example, median data may be more representative of the range of data and provide more robust predictors. However, the findings indicate that chronic and short-term findings for either the median or minimum best-fit results are all highly correlated.

Findings for BMD50 and BMDL50 have a similar pattern. As compared to the findings for BMD10 and BMDL10 results, the use of BMD50 and BMDL50 have a similar level of confidence (per the similar ranges of the Spearman and Pearson correlation results). The same is true for comparing dataset type, absolute, continuous and dichotomous are all highly correlated. Therefore, in recommending a correlation, since all are significant and highly correlated, the choice would depend mainly on the dataset type and behavior. For example, if one were analyzing continuous data, while the equations for absolute data *could* be used it would be recommended that the correlations for continuous data be selected. Both are highly correlated and therefore the most appropriate equation would be based on the type of data and not the quality of the predictive relationship.

It is important to note however that the prediction of an individual chronic value for a single chemical would involve considerable uncertainty, between 10-fold and 100-fold. This can be seen in the large spread of data in Figure 2 and Figure 3. Further, Figure 4 provides an example highlighting that prediction of a single new observation could entail a prediction error of this magnitude.

All of these findings indicate that short-term non-cancer toxicity data may provide a good indication of non-cancer chronic toxicity levels. This finding is true for continuous, dichotomous and absolute datasets, as well as between minimum or median BMD and BMDL findings, BMD50 and BMDL50, and BMD10 and BMDL10.

### 4.1 Study limitations

Limitations of this study were introduced during chemical identification and the modeling phase. In identifying significant health endpoints to be modeled as datasets in

BMDS, the assumption of this study was that those endpoints identified by the pathologist and reported in the abstract of NTP TRs were the *only* significant non-cancer findings. However, it is possible that pathologist identification, especially for the presence of non-neoplastic lesions, could have varied between NTP reports. Also, as Wang and Gray point out, possible changes in NTP practices or pathology nomenclature over time could have impacted reporting (2) although the inclusion criterion of TRs published between 2000 and 2012 was intended to minimize the chances of this occurring.

Similarly, by assessing only NTP identified chemicals it is possible that there is some pre-selection bias in how and which chemicals were selected. Generally, NTP uses a process to elect chemicals based upon the prolific use in the public or an emerging use of some chemical (34). This could mean that chemicals in this study might not be from a completely random sample. Further, a limitation of the benchmark dose approach in general is that only a limited number of animals and dose groups are tested and since this approach focuses on finding a dose corresponding to a predefined response, having limited doses and animals can be impactful (35).

Another important note is that allometic scaling for interspecies dose extrapolation was not performed on this data. Also, by using non-cancer endpoints alone it was not considered if an increase in neoplastic lesions influenced the non-cancer response.

## 4.2 Conclusions

The findings indicate that regardless of the chemical or health endpoint, short-term animal studies reasonably provide a quantitative estimate of chronic non-cancer toxicity levels: BMD10, BMDL10, BMD50 and BMDL50. This can allow for faster development of PODs for use in developing reference values like the reference doses (21) or margins of exposure  (21) or for direct comparison between chemicals like that in Chapter 4.  This finding is highly correlated, and true for continuous, dichotomous and absolute datasets, as well as between minimum or median BMD and BMDL findings.

# Chapter 3. Assessing Implicit Assumptions in Toxicity Testing Guidelines

## 1 Background

Traditional toxicology testing guidelines recommend using mice and rats of both sexes as test subjects to predict human health toxicity outcomes (2) (10) (11) (12) . This recommendation appears to maintain consistency and comparability between tests and testing guidelines. Mice and rats have significant conventional knowledge, historical availability, and ease of care and breeding, making them ideal test subjects, (36). Testing of both sexes of these rodents ensures completeness in the assessment of possible health effects and sensitivities from chemical exposures. (36) (37) (38)

The goal of toxicity testing is generally to assess human toxicity to chemicals and find the critical effect in the most sensitive sex and species to be used in chemical assessments and regulatory applications (6). Relative species or species-sex sensitivities can greatly influence regulatory toxicology determinations. Although there is not substantial literature describing *why*, the recommendation to assess all four species sex groups (male rats, female rats, male mice, and female mice) appears to be a convention adopted to ensure that the critical effect is captured. Discussions of relative sensitivities of animal subjects in the literature are generally considered within the context of a specific chemical or a specific health endpoint.

This study has operationalized the conventional recommendation of testing mice and rats of both sexes to mean that over a large group of chemicals, any one of these groups may be the most sensitive. Meaning that the distribution of species-sex sensitivities over a large group of chemicals would be a uniform distribution. This study refers to this uniform distribution as an *implicit assumption of uniform species-sex sensitivities*. The results of a large number of chronic and short-term toxicology tests for non-cancer effect levels in mice and rats of both sexes were analyzed to determine if the most sensitive species-sex distribution is uniform or if certain groups are found to be most sensitive more frequently.

This study focuses on non-cancer endpoints. Dichotomous and continuous data were considered in this study because they are a sensitive mark of toxicity and regulatory bodies such as the EPA use both types to set reference values (8).

## 1.1 Guidelines for Species-Sex Selection in Toxicity Selections

This section provides examples of notable toxicity testing and study design guidelines. All of these testing guidelines highlight examples where mice and rats of both sexes are recommended as test subjects. Generally the rationale for testing is convention, ease of care, or comparability, with less emphasis on relative sensitivities.

According to the Organization for Economic Cooperation and Development's (OECD) *Guidelines for Testing of Chemicals, Test Guideline for Chronic Toxicity*, the majority of chronic toxicity studies are carried out in rodent species and "both sexes should be used" (11). While not explicitly mentioned, the implication is to use rodents of both sexes for comparability and completeness. They state that rats and mice are "preferred models" (11) because of their short life span, their widespread use in other studies, their susceptibility to developing tumors, and the availability of strains; because of these factors there is a large amount of information available on their physiology and pathology (11).

The Canadian Council on Animal Care suggests that mice models were relied upon over the last 30 years because there are more data and best practices available for these study subjects and their genome are mapped better. Further, laboratory rodents in general are frequently perceived to have relatively few requirements other than basic housing, husbandry, and dietary needs (10).

Similar to the OECD, the NTP begins their section about "Experimental Animal Selection" in their guidance on toxicology and carcinogenesis study designs, "toxicity and carcinogenesis studies are conducted on mice and rats" (12), with minimal discussion as to *why*. The discussion of animal selection mentions which strains of mice and rats make the best test subjects (12). After a thorough review of the NTP database of TRs it can be concluded that test subjects other than mice and rats are used by NTP but it is clearly not typical practice (12) (18).

There have been notable areas in toxicology testing where differences in species sensitivities are considered. Thalidomide, a drug widely used in the 1950s and 1960s to treat morning sickness, was eventually linked to thousands of birth defects in children. However, initially the link between this drug and its developmental effects were not determined because rat experiments did not produced comparable malformations. However, after testing rabbits the link between thalidomide and the birth defects was confirmed. This episode led to the development of systematic developmental toxicology testing protocols in the U.S. and internationally which requires developmental toxicology testing to be conducted on two species, one of which cannot be a rodent. (39)

This study modeled and determined best-fit, minimum BMDL10s for a large group of chemicals and assessed the resulting distribution of species-sex sensitivities to determine the uniformity of the results.

## 2 Methods

The following section provides the methodology used for analyzing the relative sensitivity of each species-sex test group in NTP short-term and chronic toxicity tests. TRs (and Toxicity reports, where applicable) reported by the NTP between 2000 and 2012 were used. The dose-response results for all non-neoplastic endpoints as well as organ and body weight, identified as significant by the NTP pathologists, were modeled using the Environmental Protection Agency's (EPA) Benchmark Dose Software (BMDS) Version 2.6.1. BMDL10s from best-fit models for each endpoint were identified. Then the species and sex associated with the minimum BMDL10 for each chemical was identified. This section will describe this process in detail.

## 2.1 Data Collection

Data collection began by extracting publically available TRs associated with NTP chronic Toxicity and Carcinogenicity Studies from their website (18). Most TRs include summaries of major findings associated with short-term studies for the same chemical substance. In addition, Toxicity reports, which describe short-term studies in more detail, were accessed if more detailed was needed (40).

All NTP TRs, and where necessary Toxicity reports, were reviewed for specified chemical inclusion criteria, where 41 chemicals met the criteria. For a list of these chemicals refer to in the Results section. The inclusion criterion were as follows: (1) the availability of both short-term and chronic toxicological studies; (2) exposure routes had to be through one of the following methods: feed, drinking water, or gavage; (3) TR final publication date between 2000 and 2012. This time frame was chosen because it provided ample, recently published data in final publication form.

For TRs that met the above criteria, the next step was to identify datasets of interest within each TR. Within the abstract of each TR a summary of significant findings is provided as assessed by the pathologists. Only those endpoints identified in the abstract are assessed in this study because those are assumed to be the significant health effects. This study assessed non-neoplastic lesions (dichotomous endpoints) and body and organ weights (continuous endpoints). For dichotomous endpoints, a dataset within the context of this study includes the count of subjects (including species, sex and endpoint) within each exposure group where the pathologist identified a particular non-neoplastic lesions (note that severity categories were not considered). For continuous endpoints, a dataset is defined as either the mean body weight or mean organ weight for each dose group.

Doses were generally reported by NTP in mg/kg of body weight. Therefore, each dataset consisted of the following:

a. Chemical and TR (41 chemicals total): see Table 10 for or a list of chemicals that met the chemical inclusion criteria.

b. Duration (chronic and short-term)

    i. Chronic is approximately 2 years

    ii. Short-term is approximately 13 weeks.

c. Total number of animal test subjects

d. Species (F344/N and Wistar rat; B6C3$F_1$ mice)

e. Sex (male and female)

f. Dose group (for example, 0, 5, 10, 25, 55, 90 mg per kg body weight)

g. Health effect:

    i. Continuous endpoints: average final body weights; or average organ weight at the various dose group. Standard errors, which were converted into standard deviation using historical controls were also extracted.

ii. Dichotomous endpoints: Incidence of non-neoplastic lesions at each dose group

For continuous data points, the adverse direction must be recorded as this is an important modeling assumption and was not as consistent as one may assume. For example, generally body weight was observed to go down as dose increased but there are notable examples where body weight actually increased following an exposure (41). Consider female rats in TR512 (Elmiron®), it is noted in the abstract that as compared to vehicle controls, “the mean body weight of all dosed groups of female [rats in short-term testing] were greater” (41). Once the data collection was complete, each dataset was then fit with between five and eight possible models to determine the best-fit model, which is discussed in the next section.

## 2.2 Batch-Processing and Modeling in BMDS

In order to assess the uniform species-sex sensitivity assumption, first all datasets were individually modeled for best-fit model determination. Ultimately, the minimum best-fit BMDL per chemical and study duration (this combination is henceforth referred to as *chemical-duration*) could then be binned by species-sex identifiers. Then the next sections will describe the best-fit BMDL identification and the assessment of species-sex sensitivities over the entire chemical sample.

All findings determined by the NTP, even non-significant findings, are reported in excel files the Chemical Effects in Biological Systems Database (CESB) (34). Using U.S. EPA BMDS Version 2.6.1, R code and visual basic, batch processing techniques were developed to extract data from relevant endpoints from the excel files of total findings in CEBS. Then each dataset was converted into single comma separated values (csv) files and then to data files (.dax) to be utilized by BMDS. These data files were then further batch processed into session files to be modeled in BMDS. Based on the BMDS user guidance for dichotomous data the dose-response models fit to all datasets included: 1) Gamma, 2) Logistic, 3) Log-Logistic, 4) Log-Probit, 5) First, second and third degree multistage, 6) Probit, 7) Weibull, and 8) Quantal-Linear. For continuous data the dose-response curves included: 1) Exponential, 2) Hill, 3) Linear, 4) first and second degree polynomial, and 5) Power. (20)

For continuous data both modeled and constant variability were evaluated for every dataset within each chemical of interest (18). Then a series of post-modeling logical arguments based on the BMDS user-manual were applied to the results to identify whether model or constant variability was appropriate. The logical arguments applied are based on BMDS output narratives for Tests 2 and 3, which state:

> *Test 2: Tests the null hypothesis that variances are homogeneous. If this test fails to reject the null hypothesis, the simpler constant variance model may be appropriate.*
> *Test 3: Tests the null hypothesis that the variances are adequately modeled. If this test fails to reject the null hypothesis, it may be inferred that the variances have been modeled appropriately.*
> (20)

BMDS output was saved in Excel spreadsheets which were programmed to report the following information: dataset (i.e. health endpoint), model type, specified effect, BMD, BMDL, p-value (or multiple p-values for continuous data), Akaike Information Criterion (AIC), scaled residual for control and dose groups. These raw results were next assessed to identify the best fit model for each dataset.

### 2.2.1 Filters

Once all datasets were modeled in BMDS, a best-fit model and thus a best-fit BMDL were identified for each dataset. BMDS user guidance and other considerations were used to create a series of logical arguments, or *filters*, which were applied to identify best-fit models efficiently.

It is important to note that any single endpoints removed as a result of these filters was tracked, because it had to be considered if any datasets were completely removed (*i.e*. if no model adequately fit any of the endpoints for the dataset). The filters are described below.

Filter 1. If any modeling error was reported, the data for the endpoint was removed and tracked separately. This includes actual error indications, as well as physically impossible findings (e.g. BMD= -999999). Physically impossible findings are assumed to be the result of data that simply cannot be modeled by BMDS.

Filter 2. Examine the chi-square goodness-of-fit p-value and chi-squared residuals for each model generated. The p-value provides an indication of how well the model fits the observed dose-response data. The chi-squared residuals provide an indication of how well a model fits the dose group closest to the calculated BMD or BMDL. Modeled data was retained if the p-value was greater than 0.1 *and* the absolute value of the scaled residuals was less than or equal to 2.0. Any data not meeting these criteria is removed and tracked separately. (20)

Filter 3. If more than one model was found to meet the best-fit criteria, the AIC was assessed within the remaining model choices for each dataset. According to the BMDS user's guide, the AIC can be used to compare different types of models which use a similar fitting method (20). The AIC should only be used in cases of models with BMDL values within a threefold difference. (20).

Filter 4. After finding the minimum AIC some datasets still had multiple best-fit models and the same AICs. According to the BMDS user guide, at this point visual assessment and a review of literature for assessing realistic results should be implemented to determine the best fit amongst the remaining models. However, with the vast quantity of data being examined this was not practical. As a result, BMD to BMDL ratios were determined. Ringblom, *et al*. found BMD to BMDL ratios usually fell between 3 and 10 (22). As such, datasets having BMD to BMDL ratios above 10 are not typical of a biological response (22), and were removed.

Filter 5. Any datasets where there were still multiple best-fit models were then averaged together, although generally these instances resulted in BMD and BMDLs which were virtually or completely identical. (22) (23)

After the application of these five filters, each dataset had a single (or averaged) best-fit model with a predicted BMD and BMDL; recall that a *dataset* includes the dose-response values for a particular health endpoint (i.e. body or organ weight, or presence of a particular lesion) for a particular species and sex exposed to a particular chemical. For a list of the total number of *datasets* before and after the application of filters, refer to the results section Table 12.

With numerous endpoints identified for each chemical and duration, the next step was to determine the minimum, best-fit BMDL for each chemical within a study duration (*i.e*. chronic or short-term), as well as the species and sex associated with this value. Refer to the results section for reports the minimum, best-fit BMDLs for each chemical, and for the total number of datasets per chemical (i.e. the number of datasets available per chemical prior to the selection of a minimum per chemical). A minimum BMDL for each chemical and duration was determined in three ways: 1) for continuous endpoints, 2) for dichotomous endpoints, and (3) for all endpoints (referred to as the *absolute BMDL*)"This was performed to determine if the type of endpoint had influence in the sensitivity effects.

After determining the minimum, best-fit BMDL for each chemical, the species, species-sex, and endpoint type (dichotomous or continuous) of each minimum BMDL were binned by group. The next step was to apply statistical assessments to binned results to determine the significance of the findings.

## 2.3 Statistical Analysis

Once the resulting best-fit, minimum BMDLs were binned by species, species-sex, and dataset type (continuous or dichotomous), the results were assessed to determine the significance of the observed trends. The statistical test of choice for the assessment of the binned species, species-sex, and endpoint type of data is the chi-squared test for goodness-of-fit. To test the implied assumption in many toxicity testing guidelines suggest that all species sex combinations must be tested to capture the most sensitive group, the statistical analysis will focus on whether there is a uniform distribution amongst the species and species-sex groups. The null hypothesis is that the distribution among the four species-sex groups follow a uniform distribution and therefore 25% of the best-fit, minimum BMDLs per chemical-duration would fall into each species-sex group; or in the case of analyzing species group, the null hypothesis would be that 50% of the best-fit, minimum BMDL per chemical-duration would fall into each species group. For species or endpoint group the null hypothesis is that 50% of the binned results fall into the associated group (i.e. dichotomous versus continuous; rats versus mice”). The alternative hypothesis is that these values do not follow a uniform distribution. The null hypothesis is rejected at p≤ 0.05 and the findings would be that at least one group has a significantly higher chance of possessing the minimum BMDL. No pairwise comparison were performed at this step.

To further assess which species-sex group(s) differ significantly from a proportion of 25%, as assumed in the null hypothesis, a pairwise chi-squared goodness-of-fit test was performed for each of the groups individually by comparing the target group to the summation of the other three species-sex groups. Therefore the null hypothesis was tested by determining if these counts follows a ratio of 1:3 (i.e., that their respective probabilities are 0.25 and 0.75).For these comparisons the Bonferroni adjustment was utilized (32) (i.e., a p-value of 0.0125 (0.05/4)), where the null hypothesis would therefore be a uniform distribution where 25% of the results would fall into the species-sex category of interest, and the remaining 75% would fall into the sum of the other categories.

The next section discusses the results of the methodology described.

## 3 Results

This section provides the results of the investigation including the chemicals from the NTP TRs which met the chemical inclusion assessment. Also reported in this section are the number of raw datasets versus best-fit endpoints. Raw datasets correspond to the pathologist identified dose-response data used as input to BMDS and best-fit endpoints are those endpoints remaining after the modeling and filtering process. Also reported are the distributions of minimum, best-fit BMDLs across datasets, species, and species-sex categories, as well as the resulting statistical significance of these observed distributions.

### 3.1 Chemical Inclusion Results

Table 10 describes those chemicals tested by NTP that met the chemical inclusion criteria. It is important to note that while all of the chemicals listed in Table 10 met the chemical inclusion criteria three of these chemicals are not represented in the final results for one study-duration: TR512 (chronic), TR501b (short-term) and TR565). These three chemicals are represented Table 10 with an asterisk and in Table 11 by hash marks. They were excluded because: (1) no significant health effects were reported in one study duration of a particular chemical; (2) there was no best-fit model that BMDS could fit to the chemical or chemical-duration; (3) best-fit BMDLs were identified but were highly uncertain because the BMDL to BMD ratio was greater than 10. In subsequent results, chemicals are referred to by the TR.

*Table 10: NTP analyzed chemicals meeting chemical inclusion criteria.*

| Report # | Toxicity report* | Chemical | CAS | Exposure |
|---|---|---|---|---|
| TR470 | | Pyridine | 110-86-1 | Drinking Water |
| TR476 | | Primidone | 125-33-7 | Feed |
| TR491 | | Methyleugenol | 93-15-2 | Gavage |
| TR493 | | Emodin | 518-82-1 | Feed |
| TR494 | | Anthraquinone | 84-65-1 | Feed |
| TR495 | | Sodium Nitrite | 7632-00-0 | Drinking Water |
| TR497 | TOX-47 | Methacrylonitrile | 126-98-7 | Gavage |
| TR498 | TOX-23 | p-Nitrotoluene | 99-99-0 | Feed |
| TR501 | | p,p'-Dichlorodiphenyl Sulfone | 80-07-9 | Feed |
| TR504 | TOX-44 | o-Nitrotoluene | 88-72-2 | Feed |
| TR505 | | Citral | 5392-40-5 | Feed |
| TR506 | | Acrylonitrile | 107-13-1 | Gavage |
| TR508 | Tox-27 | Riddelliine | 23246-96-0 | Gavage |
| TR509 | | 2,4-Hexadienal | 142-83-6 | Gavage |

|  |  |  |  |  |
|---|---|---|---|---|
| TR510a |  | Urethane + 0%Ethanol | 51-79-6 | Drinking Water |
| TR510b* | TOX-52 | Urethane + 5%Ethanol | 51-79-6 & 64-17-5 | Drinking Water |
| TR511 |  | Dipropylene Glycol | 25265-71-8 | Drinking Water |
| TR512* |  | Elmiron® | 37319-17-8 | Gavage |
| TR514 |  | trans-Cinnamaldehyde (Microencapsulated) | 14371-10-9 | Feed |
| TR516 |  | 2-Methylimidazole | 693-98-1 | Feed |
| TR533 |  | Benzophenone | 119-61-9 | Feed |
| TR535 | TOX-67 | 4-Methylimidazole | 822-36-6 | Feed |
| TR537 |  | Dibromoacetic Acid | 631-64-1 | Drinking Water |
| TR540 |  | Methylene Blue Trihydrate | 7220-79-3 | Gavage |
| TR541 |  | Formamide | 75-12-7 | Gavage |
| TR544 |  | Dibromoacetonitrile | 3252-43-5 | Drinking Water |
| TR546 | TOX-72 | Sodium Dichromate Dihydrate | 7789-12-0 | Drinking Water |
| TR549 |  | Bromochloroacetic Acid | 5589-96-8 | Drinking Water |
| TR550 | TOX-9 | Cresols | 1319-77-3 | Feed |
| TR551 |  | Isoeugenol | 97-54-1 | Gavage |
| TR554 |  | 5-(Hydroxymethyl)-2-Furfural | 67-47-0 | Gavage |
| TR557 |  | B-Myrcene | 123-35-3 | Gavage |
| TR558 | TOX-65 | 3,3',4,4'-Tetrachloroazobenzene | 14047-09-7 | Gavage |
| TR560 |  | Androstenedione | 63-05-8 | Gavage |
| TR562 |  | Goldenseal Root Powder | GOLDENSEALRT | Feed |
| TR563 |  | Pulegone | 89-82-7 | Gavage |
| TR565* |  | Milk Thistle Extract | 84604-20-6 | Feed |
| TR570 |  | Alpha,Beta-Thujone | 76231-76-0 | Gavage |
| TR571 |  | Kava Kava Extract | 9000-38-8 | Gavage |
| TR575 |  | Acrylamide | 79-06-1 | Drinking Water |
| TR579 |  | N,N-Dimethyl-P-Toluidine | 99-97-8 | Gavage |
| * One duration of results met chemical inclusion criteria but not represented in final results. |  |  |  |  |

## 3.2 Minimum, Best-Fit BMDL per Chemical-Duration

To assess the assumption of uniform species-sex sensitivity distributions in general toxicology testing guidelines, best-fit models were determined in BMDS and resulting BMDLs were identified for each dataset using the filters described in the Methods Section. Then the minimum BMDL within a chemical-duration category and its associated species and sex were determined. These results are reported in Table 11. Note that Table 11 represents the *absolute* minimum, best-fit BMDL per chemical-duration. Minimum, best-fit BMDLs were also determined per chemical-duration for continuous datasets and dichotomous datasets separately. Refer to Table 10 for a list of chemical names and corresponding technical report numbers.

*Table 11: Absolute minimum, Best-fit BMDLs per chemical-duration.*

| | **Chronic** | | | | **Short-Term** | | | |
|---|---|---|---|---|---|---|---|---|
| **TR#** | **Min BMDL** | **Species** | **Sex** | **Dataset type*** | **Min BMDL** | **Species** | **Sex** | **Dataset type** |
| TR470 | 1.76 | Rats | m | d | 4.03 | rats | f | c |
| TR476 | 1.74 | Rats | f | c | 2.94 | rats | f | d |
| TR491 | 2.86 | Rats | m | c | 0.48 | rats | m | d |
| TR493 | 207.62 | mice | m | c | 176.44 | mice | f | c |
| TR494 | 2.76 | mice | f | d | 0.46 | mice | m | d |
| TR495 | 1.27 | Rats | f | c | 21.01 | rats | f | c |
| TR497 | 27.29 | mice | f | d | 6.09 | rats | f | c |
| TR498 | 0.0409 | Rats | m | d | 0.29 | rats | f | c |
| TR501a | 3.709 | Rats | f | d | 3.75 | rats | m | c |
| TR501b | 4.73 | rats | m | d | | | | |
| TR504 | 0.23 | rats | f | d | 1.94 | rats | m | c |
| TR505 | 2.00 | rats | f | c | 419.14 | mice | f | d |
| TR506 | 5.85 | rats | m | c | 7.036 | rats | f | d |
| TR508 | 0.078 | rats | m | c | 0.089 | rats | f | d |
| TR509 | 4.15 | rats | f | d | 37.87 | rats | m | c |
| TR510A | 29.58 | rats | f | d | 57.84 | mice | f | d |
| TR511 | 7.55 | rats | f | c | 14.44 | rats | m | c |
| TR512 | | | | | 315.18 | mice | f | d |
| TR514 | 81.01 | rats | m | c | 271.26 | rats | f | d |
| TR516 | 0.94 | rats | m | c | 38.46 | mice | f | c |
| TR533 | 0.23 | mice | f | c | 6.83 | rats | m | d |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| TR535 | 4.89 | mice | m | d | 13.10 | mice | f | d |
| TR537 | 8.52 | mice | m | c | 2.93 | mice | m | d |
| TR540 | 2.76 | rats | f | c | 0.16 | rats | f | d |
| TR541 | 16.64 | mice | m | c | 11.82 | rats | f | d |
| TR544 | 1.56 | rats | m | c | 13.19 | mice | f | d |
| TR546 | 0.35 | rats | m | c | 0.46 | mice | m | d |
| TR549 | 7.22 | mice | f | c | 13.10 | rats | f | d |
| TR550 | 1.87 | rats | f | c | 5.26 | rats | f | d |
| TR551 | 6.071 | mice | m | c | 18.10 | rats | m | d |
| TR554 | 25.47 | rats | f | d | 441.06 | rats | m | c |
| TR557 | 0.014 | mice | f | c | 0.040 | mice | m | d |
| TR558 | 0.321 | rats | m | c | 0.078 | rats | m | d |
| TR560 | 0.26 | rats | f | c | 0.57 | rats | f | d |
| TR562 | 70.22 | rats | f | c | 117.75 | rats | m | d |
| TR563 | 0.59 | rats | f | c | 3.91 | mice | m | d |
| TR565 | 443.16 | rats | m | c | | | | |
| TR570 | 4.20 | mice | f | d | 0.41 | rats | f | c |
| TR571 | 0.092 | rats | m | d | 0.14 | rats | m | c |
| TR575 | 0.084 | rats | m | d | 17.44 | mice | f | c |
| TR579 | 0.35 | rats | m | c | 0.0019 | rats | m | d |
| *d= Dichotomous; c= Continuous | | | | | | | | |

Prior to the application of filters there were a total of 1706 datasets from 41 chemicals included in the study. Following the application of filters there was a reduction to 1181 datasets. Datasets that are not included following the application of filters can mainly be attributed to an inability of BMDS to fit a model to the dataset or modeling error. Table 12 below describes these results in more detail.

*Table 12: Total number of datasets before and after the application of filters.*

| | |
|---|---|
| **Raw datasets (continuous)** | 399 |
| **Raw Datasets (dichotomous)** | 1307 |
| **Total raw datasets** | **1706** |
| | |
| **Best-fit models per endpoint (continuous)** | 220 |
| **Best-fit models per endpoint (dichotomous)** | 961 |
| **Total best-fit models for all endpoints** | **1181** |
| | |
| **Percent raw datasets removed** | **31%** |

### 3.3 Minimum, Best-Fit BMDL Distribution by Species-Sex

This section presents the resulting distributions of the binned minimum, best-fit BMDL10s identified per species, and species-sex groups. Goodness-of-fit and pairwise analysis results are presented for species-sex findings as well as goodness-of-fit comparisons for the species findings. Absolute minimum, dichotomous minimum, and continuous minimum datasets are presented.

Table 13 through Table 15 represent the distribution of results and the goodness-of-fit examination on the best-fit, minimum BMDLs per species-sex group associated with absolute, continuous and dichotomous findings (absolute endpoints reported in Table 11). Note that the tables are presented in terms of species for conciseness but that the actual distribution assessed was that of species-sex. In all tables, a chi-square test for goodness-of-fit was made in each row. The results indicate that the distribution of species-sex sensitivities, over the 41 chemicals and two durations, may not be uniform (p=0.05). Note that dichotomous and continuous endpoints were analyzed separately (and then combined for absolute findings), which allowed for the possible evaluation of differences in sensitivity of the two types of endpoints (*i.e.*, body weight or organ weight changes versus the presence of specific non-neoplastic lesions).

*Table 13: Absolute minimum best-fit BMDL goodness-of-fit results, species-sex*

| | | Rats | Mice | p-value | Total |
|---|---|---|---|---|---|
| Chronic* | Number | 29.00 | 11.00 | 0.042 | 40.00 |
| | Percent | 72.50 | 27.50 | 0.0044 | 100.00 |
| | | Rat | Mice | | Total |
| Short | Number | 27.00 | 19.00 | 0.145 | 46.00 |
| | Percent | 58.70 | 41.30 | 0.24 | 100.00 |

*p<0.05

*Table 14: Dichotomous minimum best-fit BMDL goodness-of-fit, species-sex*

| | | Rats | Mice | p-value | Total |
|---|---|---|---|---|---|
| Chronic | Number | 21.00 | 17.00 | 0.112 | 38.00 |
| | Percent | 55.26 | 44.74 | 0.516 | 100.00 |
| | | Rats | Mice | | Total |
| Short* | Number | 26.00 | 9.00 | 0.008 | 35.00 |
| | Percent | 74.30 | 25.71 | 0.0041 | 100.00 |

*p<0.05

*Table 15: Continuous minimum best-fit BMDL goodness-of-fit, species-sex*

| Chronic | | Rats | Mice | p-value | Total |
|---|---|---|---|---|---|
| | Number | 22.00 | 11.00 | 0.055 | 33.00 |
| | Percent | 66.67 | 33.33 | 0.056 | 100.00 |
| Short | | Rats | Mice | | Total |
| | Number | 23.00 | 13.00 | 0.343 | 36.00 |
| | Percent | 63.90 | 36.11 | 0.096 | 100.00 |

*p<0.05

Table 16 through Table 18 represent the results for pairwise comparison of the binned best-fit, minimum BMDLs per species-sex group associated with absolute, continuous and dichotomous findings. In each table, chi-square tests for equality of numbers. The results indicate that the distribution of species-sex sensitivities, over the 41 chemicals and two durations, may not always be uniform. It is statistically significant (p= 0.00125) in a few instances, particularly continuous endpoints where male rats are apparently more sensitive than the other species-sex groups, principally for chronic continuous endpoints and dichotomous short-term endpoints.

Note that there are individual p-values for *each* species-sex group reported. The p-values represent the test of goodness-of-fit of 25% versus 75% between the target group and the sum of the other groups. As stated, the significance was determined at a level of p= 0.0125, therefore, based on the below results no significant difference was detected in Table 16. But significant examples are represented in Table 17 and Table 18.

*Table 16: Absolute minimum best-fit BMDL pairwise comparison, species-sex*

| | | **Rats** | | **Mice** | | |
|---|---|---|---|---|---|---|
| | | **Male** | **Female** | **Male** | **Female** | **Total** |
| **Chronic** | **Number** | 15.00 | 14.00 | 5.00 | 6.00 | 40.00 |
| | **P-value** | 0.067 | 0.14 | 0.068 | 0.14 | |
| | **Percent** | 37.50 | 35.00 | 12.50 | 15.00 | 100.00 |
| | | **Male** | **Female** | **Male** | **Female** | **Total** |
| **Short** | **Number** | 13.00 | 14.00 | 5.00 | 8.00 | 46.00 |
| | **P-value** | 0.27 | 0.14 | 0.068 | 0.47 | |
| | **Percent** | 28.26 | 30.43 | 10.86 | 17.39 | 100.00 |

*p<0.05 **p<0.0125

*Table 17: Dichotomous minimum best-fit BMDL pairwise comparison, species-sex*

| | | **Rats** | | **Mice** | | |
|---|---|---|---|---|---|---|
| **Chronic** | | **Male** | **Female** | **Male** | **Female** | **Total** |
| | **Number** | 6.00 | 15.00* | 6.00 | 11.00 | 38.00 |
| | **p-value** | 0.19 | 0.039 | 0.19 | 0.57 | |
| | **Percent** | 15.79 | 39.47 | 15.79 | 28.95 | 100.00 |
| **Short** | | **Male** | **Female** | **Male** | **Female** | **Total** |
| | **Number** | 16.00** | 10.00 | 2.00* | 7.00 | 35.00 |
| | **p-value** | 0.0047 | 0.63 | 0.0084 | 0.49 | |
| | **Percent** | 45.71 | 28.57 | 5.71 | 20.00 | 100.00 |

*p<0.05 **p<0.0125

*Table 18: Continuous minimum best-fit BMDL pairwise comparison species-sex*

| | | Rats | | Mice | | |
|---|---|---|---|---|---|---|
| | | **Male** | **Female** | **Male** | **Female** | **Total** |
| **Chronic** | **Number** | 15.00** | 7.00 | 5.00 | 6.00 | 33.00 |
| | **P-value** | 0.0067 | 0.62 | 0.19 | 0.37 | |
| | **Percent** | 45.45 | 21.21 | 15.15 | 18.18 | 100.00 |
| **Short** | | **Male** | **Female** | **Male** | **Female** | **Total** |
| | **Number** | 10.00 | 13.00 | 7.00 | 6.00 | 36.00 |
| | **P-value** | 0.70 | 0.12 | 0.44 | 0.25 | |
| | **Percent** | 27.78 | 36.11 | 19.44 | 16.67 | 100.00 |

*p<0.05 **p<0.0125

Table 13 thought Table 15 represent the results for the goodness-of-fit comparison of the binned best-fit, minimum BMDLs per species group associated with absolute, continuous and dichotomous findings. The results indicate that the distribution of species sensitivities, over the 41 chemicals and two durations, may not be uniform. It is statistically significant (p= 0.05) in a few instances, particular amongst the absolute findings.

*Table 19: Absolute minimum best-fit BMDL pairwise comparison, species*

| | | Rats | Mice | p-value | Total |
|---|---|---|---|---|---|
| **Chronic** | **Number** | 29.00* | 11.00 | 0.0044 | 40.00 |
| | **Percent** | 72.50 | 27.50 | | 100.00 |
| | | | | | |
| **Short** | **Number** | 27.00* | 19.00 | 0.026 | 46.00 |
| | **Percent** | 58.70 | 28.26 | | 100 |

*p<0.05

*Table 20: Dichotomous minimum best-fit BMDL pairwise comparison, species*

| | | Rats | Mice | p-value | Total |
|---|---|---|---|---|---|
| **Chronic** | **Number** | 21.00 | 17.00 | 0.52 | 38.00 |
| | **Percent** | 55.26 | 44.74 | | 100 |
| | | | | | |
| **Short** | **Number** | 26.00* | 9.00 | 0.0041 | 35.00 |
| | **Percent** | 74.30 | 25.71 | | 100.00 |

*p<0.05

*Table 21: Continuous minimum best-fit BMDL pairwise comparison, species*

| | | Rats | Mice | p-value | Total |
|---|---|---|---|---|---|
| **Chronic** | **Number** | 22.00 | 11.00 | 0.055 | 33 |
| | **Percent** | 66.67 | 33.33 | | 100.00 |
| | | | | | |
| **Short** | **Number** | 23.00 | 13.00 | 0.34 | 36.00 |
| | **Percent** | 53.49 | 46.51 | | 100.00 |

## 4 Discussion

This study assessed the relative toxicity distribution of mice and rats of both sexes over a large group of NTP toxicity testing results, reported in their TRs. To investigate this, best-fit minimum BMDLs for chronic and short-term toxicity testing results of 41 chemicals of varied application were modeled using BMDS. The resulting species-sex distributions were assessed. Findings indicate that for the group of chemicals assessed in this study that there is not a uniform species or species-sensitivity distribution. Indicated in Table 13 through Table 15, for this set of chemicals, the goodness-of-fit results suggest that rats tend to be significantly more sensitive than mice. Further, indicated in Table 16 through Table 21, the pairwise comparison results suggest that male rats are more sensitive than the sum of the other groups for chronic continuous and short-term dichotomous endpoints.

All of these findings are considered within the context of dose, instead of the traditional context of health effect, and may have important implications for toxicity testing guidelines. The results raise questions about the need to test all four species-sex groups. Improving upon this could increase efficiency in toxicity testing and reduction of resource burdens.

### 4.1 Study Limitations

Limitations of this study were introduced during chemical identification and modeling. In identifying significant health endpoints to be modeled as datasets in BMDS, the assumption of this study was that those endpoints identified by the pathologist and reported in the abstract of NTP TRs were the *only* significant findings. However, it is possible that pathologist identification, especially for the presence of non-neoplastic lesions, could have varied between NTP reports. Also, as Wang and Gray point out, possible changes in NTP practices or pathology nomenclature over time could have impacted reporting (2).

Similarly, by assessing only NTP identified chemicals it is possible that there is some pre-selection bias in how and which chemicals were selected. Generally, NTP uses a process to elect chemicals based upon the prolific use in the public or an emerging use of some chemical (34). This could mean that chemicals in this study might not be from a completely random sample. Further, the limitation of the benchmark dose approach is that only a small number of animals and dose groups are tested. Since this approach focuses on finding a dose corresponding to a predefined response, having limited doses and animals can be impactful to the power of the study (35).

Another important note is that allometic scaling for interspecies dose extrapolation were not performed on this data, although since the scaling would have likely made rats appear more sensitive, the results reported in this study would likely be amplified by scaling. Further, Wistar rats and F344/N rat strains were combined, which could have masked certain findings. However, since mice were a single strain there is added reliability of those findings. Also, by using non-cancerous endpoints alone it was not considered if an increase in neoplastic lesions influenced the non-cancer response.

## 4.2 Conclusions

This study provides a strong indication that species and species-sex sensitivities may not always be uniform. For the set of chemicals considered, rats were more sensitive than mice and male rats were more sensitive than the combination of the other groups in a few instances (depending on the endpoint type). The sheer amount of data processed and modeled in this study is an asset to this finding and unlike any other studies of which the authors are aware. These findings could have implications for future toxicity testing guidelines and encourage more rigorous consideration of the relative toxicity within toxicity testing guidelines.

# Chapter 4. Comparing Human Health Toxicity of Alternative Chemicals

## 1 Introduction

The goal of this study is to demonstrate methods to consider relative chronic toxicity in determining the desirability of chemical alternatives, when the alternative chemical lacks chronic toxicity data but the target does not. Massachusetts Institute of Technology (MIT) Chemical Purchasing Wizard was selected as a prototypical chemical alternative identification tool because it has clearly identified alternative recommendations for specific target chemicals in specific functions. It must be emphasized that this paper is not singling out MIT's framework, but merely selected this tool for its ease of use and representativeness. This study used predictive relationships developed by Kratchman *et al*. (refer to Chapter 2, Equation 1) and short-term toxicity data to predict chronic toxicity outcomes for proposed chemical alternatives that lack chronic toxicity data. These findings were then compared to experimentally derived chronic BMDL10s for the associated target chemicals and a human toxicity judgment for the pair was determined.

As the public demands safer chemical alternatives in products and processes there is a need for ways to compare the relative human toxicity of target and alternative chemicals. In addition to assessing health and toxic effects, alternative assessments often introduce environmental objectives including other values as "legitimacy, equity, public engagement, and accountability" (42) . There is ongoing debate about how to identify "safer" alternatives, and it is possible that different approaches may yield different chemical choices (4) (42).

A specific area of debate is the role of relative toxicity of chemicals in judging the desirability of alternatives (43). Concern about potential harm to human health or the environment is a key element in identifying chemicals of concern as targets for substitution. Yet these data may be lacking for alternative chemicals because hundreds of new chemicals (and potential substitutes) are being synthesized every year (7) and existing chemicals have different amounts of toxicological data. This has the potential to result in substitutions based on an imbalanced understanding of the target and alternative chemicals; this study is specifically focused situations in which the alternative has little or no chronic toxicity data, but the target chemical has well studied chronic non-cancer health effect levels.

Such disparities in data availability can result in *regrettable substitutions*. A regrettable substitution is when hazardous chemicals are replaced with "substitute chemicals or redesigned products or processes that may pose new and potentially greater hazards" (44). A lack of toxicity for a chemical can be perceived as a finding of no adverse health consequences, and therefore erroneously interpreted as safety, resulting in a regrettable substitution (45). We focus on chronic toxicity data because these are by far the most common basis for development of chronic human health reference values (13).

We assume that in assessing alternative-target chemical pairs, it is important to have some measure of relative toxicity. Traditionally, these would be in the form of chronic human health reference values (HHRVs) based upon experimentally derived chronic toxicity data. However, obtaining such data is time and resource intensive and too few chemicals have been evaluated in this manner (45), leaving chemicals in the environment without fully characterized health effects or levels, or even the erroneous assumption of safety (7) (13) (45). In addition, the National Academy of Sciences (NAS) recommends that a comparative exposure assessment be performed as part of an alternatives assessment, even in a qualitative form (43). The NAS committee notes that many existing chemical substitution frameworks focus on reducing hazards and only peripherally consider relative exposure. Recognizing that physical-chemical properties of alternatives could lead to higher or lower levels of exposure than the chemical of concern, NAS recommends assessing "whether the expected exposures from the chemical of concern and the alternatives would be substantially equivalent" (43) and if they are not to that a detailed exposure assessment might become necessary. (43)

The purpose of this study is to demonstrate methods to assess the relative toxicity of suggested chemical alternatives specifically when the target chemical possesses experimentally derived toxicity data but the alternative does not. This approach could be used if neither chemical had chronic toxicity data as well. We rely on relationships for predicting chronic non-cancer toxicity levels using short-term toxicity data developed by Kratchman *et al*. (Chapter 2). Recall that the correlations developed by Kratchman, *et al*. apply only to non-cancer endpoints. This could be a limitation in determining regrettable substitutions, especially in instances where cancerous endpoints were the most critical effect (2). However, assessing endpoints in a quantitative manner in a step forward for many chemical alternative comparisons.

Using these relationships (Chapter 2), two different target chemicals and their suggested alternatives (referred to as *target-alternative pairs*) proposed by the Massachusetts Institute of Technology (MIT) Green Chemical Purchasing Wizard (referred to as *MIT*), were assessed (46). It should be emphasized that MIT's framework was chosen because it appears to be a prototypical alternatives assessment tool. They clearly identify target and alternative chemicals for specific functions. Their stated goal is to reduce hazard potential, likely from a human health perspective, and burden to the environment (46). These recommendations range from comparisons of cancerous effects to ecological endpoints and typically lack quantitative comparison.

Through a decision framework approach, two chemical alternatives were systematically identified that simultaneously lacked chronic toxicity data but possessed standardized short-term toxicity data like that used in developing the relationships in Kratchman *et al*. (Chapter 2). A larger pool of pairs would have been ideal but a lack of data, both short-term and chronic, drove the number of pairs assessed in this study. Both of the identified alternative chemicals are recommended by MIT as replacements for target chemicals that have chronic toxicity data that is reasonably attainable (47) (48). Importantly, since these chemicals are proposed within the MIT framework as alternatives for the same function, and have relatively similar physical-chemical properties, it is assumed that the relative expose between these target-chemical pairs would be comparable (46).

The specific substitutions analyzed are d-Limonene as a substitute for toluene and MTBE as a substitute for chloroform (46) (47) (48) (49) (50). The relative toxicity of the target and substitute chemicals were assessed first by identifying and then modeling the alternative's short-term toxicity data in the Environmental Protection Agency's (EPA's) Benchmark Dose Software (BMDS), Version 2.6.1 (20) (21). The resulting short-term BMDL10s was then input into the Kratchman, *et al*., relationships to obtain the predicted chronic non-cancer BMDL10s for the alternatives (51). These values were then compared to the known experimentally derived BMDL10s of the target chemicals proposed by MIT. The assumption is, assuming exposure are relatively similar, the chemical with the higher BMDL10 would be a better choice for a particular application from a human health perspective.

While a variety of chemical substitution frameworks exist to compare chemical alternatives. MIT's was selected because it was representative of many of the available chemical substitution tools, and MIT provides a clear identification of a target and alternative chemical for a specific function, which was necessary for this assessment. It is also recognized that a chemical substitution decision may need to consider many factors in addition to human chronic toxicity (43).

## 2 Background

Many chemical substitution frameworks are available, some have been developed as voluntary initiatives or for consumer awareness while others have certain regulatory requirements. Jacobs, *et al*., performed a substantive review of similarities and differences between many of these tools noting the rising importance a result of "increasing pressures for hazardous chemical replacement" (4). Their analysis considered hazard assessment, exposure characterization, life-cycle impacts, technical feasibility evaluation, economic feasibility and decision making (4). They found that the frameworks were generally consistent in terms of process, but often addressed different endpoints, and had methodological gaps. They suggested the need for more consistency in methods and metrics (4).

Clearly asymmetry in the availability of chronic toxicity data hinders the ability to use this attribute in comparing chemicals. A number of approaches have been proposed to help inform decisions about data-sparse chemicals. For example, the work of Pennington, *et al*. proposes techniques for the evaluation and screening of non-cancer toxicology effects by deriving $\beta_{ED10}$ slope factors from bioassay and determining possibilities for "extrapolation from other more readily available measures" (16) such as no observed or lowest observed adverse effect levels (NOAELs and LOAELs).

Another approach is the development of predictive models presented by the quantitative structure-activity relationship (QSAR) methodology. This approach assumes there is a correlation between biological and structural parameters, and bioactivity within a group of chemicals (17). *Toxicity Testing in the 21^st Century* (Tox21) is a U.S. federal government approach to developing methods to characterize non-cancer health risk from chemicals without chronic animal bioassays using high throughput models. It is intended for use in prioritizing animal testing or for use in risk assessments that lack available chronic toxicity data (15) and its direct application to chemical substitution decisions awaits development.

Kratchman, *et al*., proposed that by focusing on the dose-response relationship instead of a critical health effect, that short-term data could be used to reasonably predict chronic non-cancer health effect levels. Therefore they considered the relationship between the doses associated with observed non-cancer chronic and short-term BMDL10s in sampled chemicals and developed predictive relationships. These equations were utilized by this study to assess the pairs of target and alternative chemicals.

The goal of this study is to demonstrate methods to include consideration of relative chronic toxicity in determining the desirability of chemical alternatives, even when chemicals lack chronic toxicity data. MIT's Chemical Purchasing Wizard was selected as a prototypical chemical alternative identification tool, because it has clearly identified alternative chemicals for specific target chemicals. The methodology used to assess these chemical pairs is discussed in the next section.

## 3 Methods

This section describes the methodology used to identify, assess and compare the relative chronic toxicity of alternative-target chemical pairs. A decision tree style approach was developed to systematically eliminate target-alternative chemical pairs as study subjects until only those remained that could effectively be analyzed. In order to be analyzed the alternative needed high quality and consistent short-term toxicity data like that used in developing the relationships in Kratchman *et al.* (Chapter 2). This strategy was used to reduce selection bias. The identified alternative chemical's short-term toxicity data was modeled in BMDS and a minimum, best-fit short-term BMDL10 identified. This value was then used to predict the chronic oral BMDL10 (51). Then the chronic oral BMDL10 for the target and alternative could then be compared. The underlying assumption is that, all other things being equal, the chemical with the higher BMDL10 is "safer" from a chronic non-cancer human health perspective.

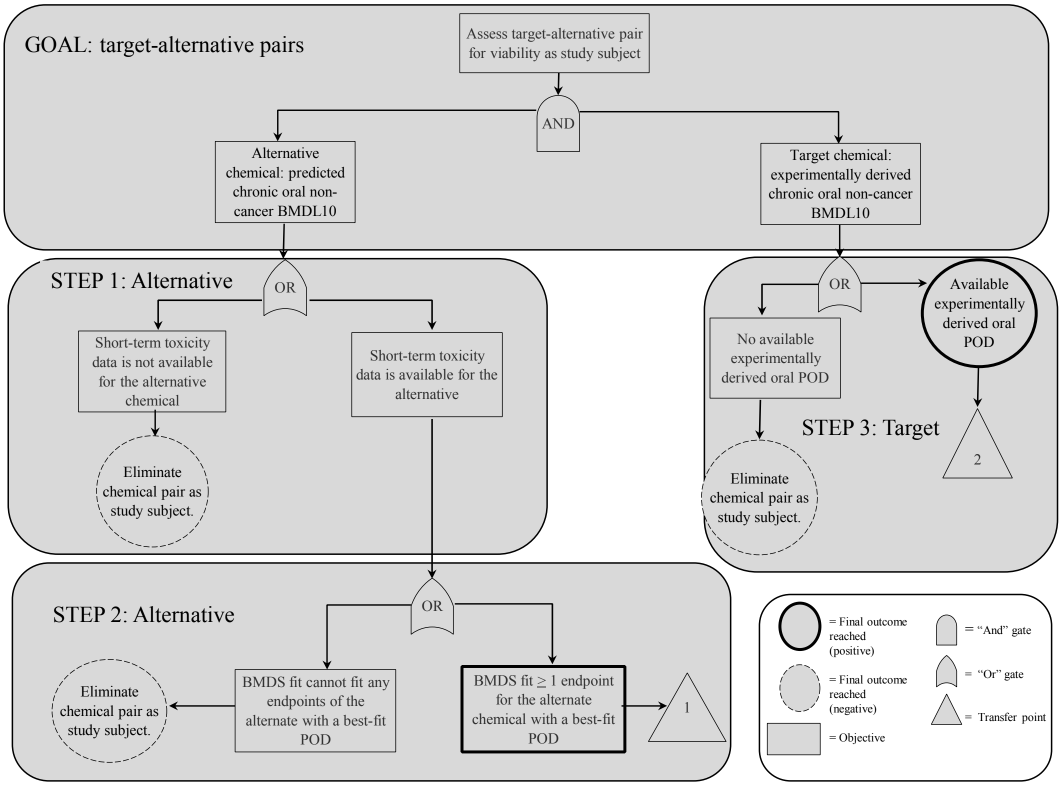
The predictive relationships developed by Kratchman, *et al*., examined the relationship between the short- and long-term toxicity data of 41 chemicals tested by the National Toxicology Program (NTP). They used all pathologist identified significant, non-cancer health. All such endpoints (also referred to as *datasets*) were modeled in BMDS and fit with best-fit models. The minimum best-fit BMDL10 was selected for each of the 41 chemicals and utilized to develop equations where short-term toxicity data can be used to predict doses associated with long-term toxicity levels. Once a BMDL10 is established for the short-term data a chronic BMDL10 can therefore be established.

## 3.1 Identification of Alternative-Target Chemical Study Subjects and PODs

This section discusses the decision framework used to systematically identify target-alternative chemical pairs. There were a total of 88 possible substitutions suggested within the MIT framework. Target-alterative pairs of interest in the MIT Chemical Purchasing Wizard were those where the target chemical has an experimentally derived oral non-cancer BMDL10s that can be found or derived through a reasonable review of the literature, but the associated alternative has only short-term toxicity data reasonably available. In addition, BMDS had to be able to fit at least one of the endpoints of the short-term data with a best-fit BMDL10, using the guidelines provided in the BMDS technical guidance (21). Each pair also was qualitatively assessed for relative exposure to ensure that exposures are generally equivalent. For each chemical listed as either a target or alternative MIT details the uses and processes for each with citations. Therefore, qualitative assessment consisted of ensuring that the usage of each target and alternative were comparable based upon the criteria and references provided. (46)

This process was operationalized in a decision-tree style approach, which is presented in Figure 5 and discussed in detail. Target-alternative pairs remaining at the end of the decision tree were considered as study subjects, and are represented by transfer points in the decision-tree (triangles). Prior to the assessment of the target-alternative pairs, a complete list of MIT proposed alternative chemicals was obtained from the website and the Chemical Abstract Services Registry Number (CASRN) was identified for each. If the CASRN of the alternative was not readily available on the American Chemical Society or EPA websites then the pair of chemicals was removed prior to entry into the decision-tree (52) (53).

*Figure 5 Target-alternative pair identification.*


GOAL: target-alternative pairs
Assess target-alternative pair for viability as study subject
AND
Alternative chemical: predicted chronic oral non-cancer BMDL10
Target chemical: experimentally derived chronic oral non-cancer BMDL10
STEP 1: Alternative
OR
Short-term toxicity data is not available for the alternative chemical
Short-term toxicity data is available for the alternative
Eliminate chemical pair as study subject.
OR
No available experimentally derived oral POD
Available experimentally derived oral POD
STEP 3: Target
2
Eliminate chemical pair as study subject.
STEP 2: Alternative
OR
Eliminate chemical pair as study subject.
BMDS fit cannot fit any endpoints of the alternate with a best-fit POD
BMDS fit ≥ 1 endpoint for the alternate chemical with a best-fit POD
1
= Final outcome reached (positive)
= Final outcome reached (negative)
= Objective
= "And" gate
= "Or" gate
= Transfer point

STEP 1. Alternative: short-term toxicity data availability

Reasonable literature reviews were conducted for all alternatives to determine if short-term data was available. Since the predictive relationships utilized were based on NTP data, NTP was considered first as a potential data source. Any alternatives that lacked NTP data were then searched using *Google Scholar* (54) with the following search criteria: "[Chemical X] short-term toxicity data", "[Chemical X] 90 day toxicity data", "[Chemical X] 13 week toxicity data", and if no data was obtained, these searches were repeated with the CASRN or synonymous chemical names. Any alternative chemicals that did not have short-term data either through NTP or from another source, was eliminated as a study subject. If an alternative were to identify that had multiple primary studies, the study with the highest quality data would be chosen. Further, per the assumptions in the predicative relationships implemented in latter steps, the test subject must have been either mice or rats and the exposure route either food, water or gavage (referred to as *oral* throughout) (51).

STEP 2. BMDS assessment of alternative chemical

For alternative-target pairs meeting the first four criteria, the alternative chemicals were then modeled in BMDS, Version 2.6.1. For dichotomous endpoints, a dataset includes the total count of subjects per exposure group and the number within each exposure group where the pathologist identified a particular non-neoplastic lesion; for continuous endpoints a dataset is defined as either the mean BW or OW for each dose group. Doses were generally reported in mg/kg of BW. (20)

Based on the BMDS user guidance for dichotomous data the dose-response models fit to all datasets included: 1) Gamma, 2) Logistic, 3) Log-Logistic, 4) Log-Probit, 5) Multistage, 6) Probit, 7) Weibull, and 8) Quantal-Linear. For continuous data the dose-response curves included: 1) Exponential, 2) Hill, 3) Linear, 4) Polynomial, and 5) Power (20). For continuous data, both model and constant variability were modeled for every dataset within each chemical of interest (18). Then, for continuous datasets, logical arguments based on the BMDS user-manual were applied to the results to identify whether model or constant variability was appropriate. The logical arguments applied are based on BMDS user guidance Tests 2 and 3, which states:

> *Test 2: Tests the null hypothesis that variances are homogeneous. If this test fails to reject the null hypothesis, the simpler constant variance model may be appropriate.*
>
> *Test 3: Tests the null hypothesis that the variances are adequately*

*modeled. If this test fails to reject the null hypothesis, it may be inferred that the variances have been modeled appropriately.*

(21)

Next the global p-value and chi-squared residuals for each dataset were assessed. The p-value provides an indication of how well the model fits the observed dose-response data. The chi-squared residuals provide an indication of how well a model fits the dose group closest to the calculated BMD or BMDL. Modeled data was retained if the p-value was greater than 0.1 *and* the absolute value of the chi-squared for residuals was less than or equal to 2.0 (21). Any dataset not meeting these criteria is removed and tracked separately. After all of the datasets for an alternative chemical were modeled at least one endpoint must have been fitted by BMDS or the alternative-chemical pair was eliminated.

The next step is to determine if the alternative chemicals identified have an experimentally determined target chemical.

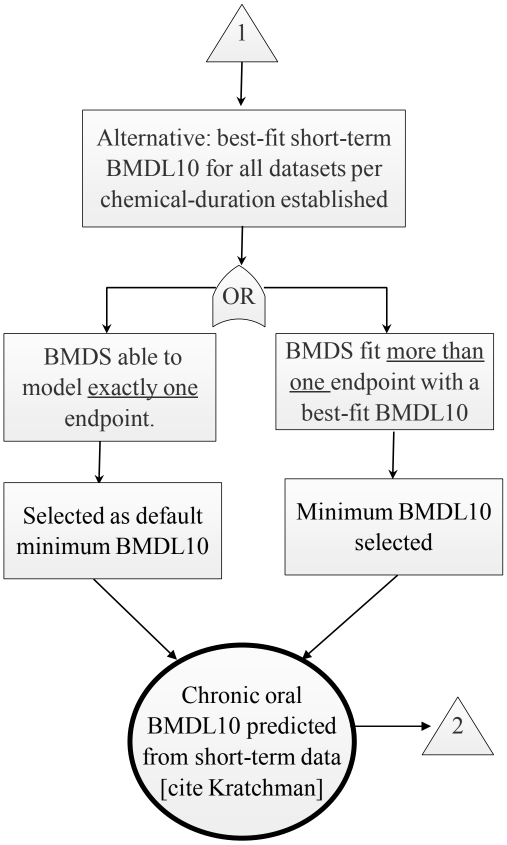
STEP 3. Target: Assess availability of experimentally derived toxicity data

For those alterative chemicals with short-term toxicity data, the target chemical must have an experimentally established oral non-cancer POD (ideally chronic data). If not, the target-alternative pair was removed as a possible study subject. Target chemicals were searched in the IRIS database of chemicals by chemical name and CASRN. If no usable POD (or data) was identified through IRIS, NTP was searched. Those chemicals still without a chronic oral POD were search through *Google Scholar* for "[Chemical X] chronic point of departure", "[Chemical X] chronic oral reference dose", "[Chemical X] long-term toxicity", and "[Chemical X] two year toxicity". Searches were repeated with CASRNs and synonymous chemical names.

## 3.2 Alternative chemical: Prediction of Chronic Oral BMDL10

In STEP 2 of Section 3.1 short-term alternative data were modeled in BMDS to determine if BMDS could fit at least one of the significant health endpoints with best-fit BMDL, those not meeting this criteria were eliminated. Now, using the same BMDS output, a short-term BMDL10 can be identified for each alternative chemical accepted into the study. Transfer point 1 from Figure 5 above is analyzed in Figure 6.

*Figure 6: Alternative chemical: predicting minimum best-fit chronic BMDL10*


1
Alternative: best-fit short-term BMDL10 for all datasets per chemical-duration established
OR
BMDS able to model exactly one endpoint.
BMDS fit more than one endpoint with a best-fit BMDL10
Selected as default minimum BMDL10
Minimum BMDL10 selected
Chronic oral BMDL10 predicted from short-term data [cite Kratchman]
2

Entering the decision pathway identified by Transfer Point 1 Figure 5 indicates that the alternative chemical's short-term toxicity non-cancer data successfully resulted in at least one best-fit BMDL10. Therefore next step is to identify the short-term BMDL10 for the alternative chemical, which was completed through the following scenarios:

1) BMDS fit *exactly one* health endpoint with a best-fit BMDL10. Although the decision pathway would remain the same, this scenario could occur in the following situations:
   a) The short-term toxicity data only had one non-cancer health endpoint available for modeling.
   b) The short-term data had multiple significant health endpoints but BMDS could only model one of the health endpoints with a best-fit model. This could be because there was a modeling error or because the global p-value or chi-squared for residuals did not meet the criteria established for a good fit (see STEP 2 of Section 3.1).
2) BMDS fit *more than one* health endpoint with a best-fit BMDL10.

The relationships later applied to predict chronic BMDL10 for the alternative chemical, requires a minimum, best-fit BMDL10 per chemical. If an alternative chemical resulted in only one short-term BMDL10, that value was treated as the minimum best-fit BMDL10. However, in 2), multiple endpoints were fit with best-fit BMDL10s, the minimum value was selected. This is consistent with the assumptions of Kratchman, *et al.* (51)

Next the alternative chemical's minimum best-fit BMDL10 short-term, was used to predict the chronic BMDL10 using the following equation:

$$y = 10^{(-0.19+0.96\times log(x))} \qquad \textit{Equation 1}$$

*Where:*
*x= short-term minimum, best-fit BMDL10 (alternative chemical)*
*y= predicted, chronic BMDL10 (alternative chemical*

(51)

The short-term BMDL10 for the alternative was assessed using Equation 1. Next the chronic BMDL10 predicted for the alternative and the experimentally derived BMDL10 for the target can be compared.

## 3.3 Comparing Target-Alternative Chronic Oral BMDL10s

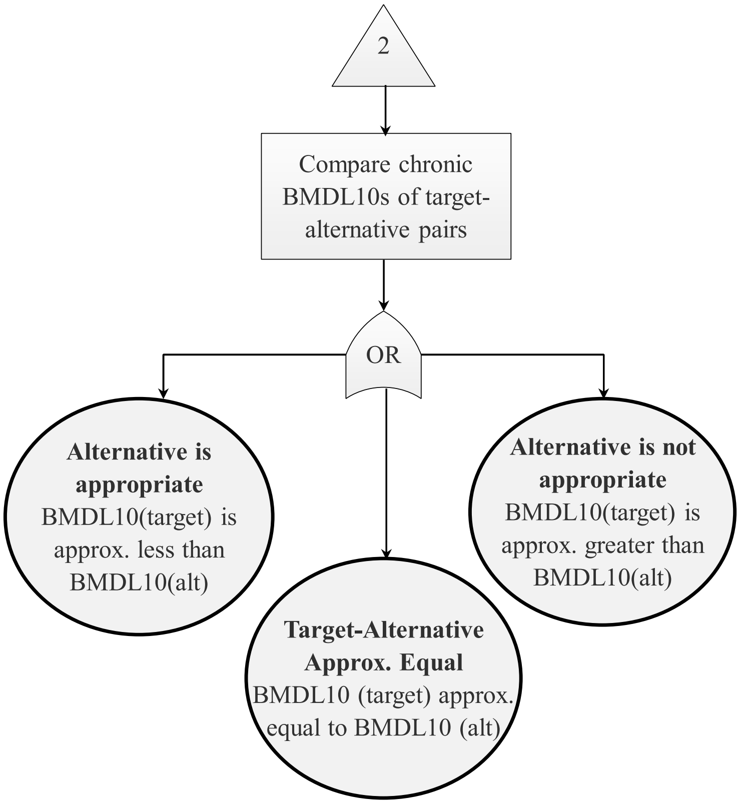
Entering the decision pathway identified by Transfer Point 1 in Figure 5 indicates that the target chemical has an experimentally derived BMDL10, and that a predicted chronic BMDL10 was established for the alternative. The next step is comparing these BMDL10 values for the target-alternative pairs to determine if the MIT recommended substitution was appropriate. The three possible outcomes of this comparison, represented in Figure 7, are as follows:

Outcome 1 Alternative is appropriate: the chronic BMDL10 for the target is approximately below the alternative chemical's chronic BMDL10. This indicates that MIT's suggested substitution is appropriate.

Outcome 2 Target and alternative are approximately equivalent: the chronic BMDL10 is approximately similar to the alternative's chronic BMDL10.

Outcome 3 Alternative is not appropriate: the target chemical's chronic BMDL10 is approximately above the alternative chemical's chronic BMDL10. This indicates that MIT's suggested substitution is not appropriate as a replacement for the proposed target. This outcome is most analogous to a regrettable substitution.

*Figure 7: Transfer point 2, comparing chronic BMDL10s of target-alternative pair*


2
Compare chronic BMDL10s of target-alternative pairs
OR
**Alternative is appropriate** BMDL10(target) is approx. less than BMDL10(alt)
**Alternative is not appropriate** BMDL10(target) is approx. greater than BMDL10(alt)
**Target-Alternative Approx. Equal** BMDL10 (target) approx. equal to BMDL10 (alt)

The next section presents the results of two target-chemical pairs identified and assessed through the decision-trees presented in this section.

## 4 Results

This section presents the results of the study where predicted chronic BMDL10s for alternatives were compared to known PODs for target chemicals. The results are presented for each step outlined in the methods section.

### 4.1 Assessment of alternatives

Recall STEP 1 was to assess the alternative chemical list provided by MIT and determine the availability of short-term non-cancer toxicity data. A *reasonable* search was conducted to identify short-term toxicity data for the list of proposed MIT alternative chemicals. This was not intended to be an exhaustive search but representative of an assessment a realistically informed user may undertake. MIT presented a total of 88 alternative chemicals; CASRNs were reasonably identifiable in 32 of these chemicals. Of these 32 chemicals, NTP short-term data was available for dichloromethane, methylbenzene, methylene chloride (OCH), and d-Limonene (18). However, dichloromethane, methylbenzene, OCH did not use an oral exposure route, so these were eliminated. Of those remaining chemicals that had CASRNs but did *not* have NTP short-term data methyl tert-butyl ether (MTBE) had an alternate source of short-term toxicity data (47). None of these alternative chemicals had multiple primary sources of short-term data that could be reasonably identified.

Therefore at the conclusion of STEP 1 the following alternative chemicals were potential study subjects: d-Limonene and MTBE. D-Limonene short-term toxicity data was obtained from NTP, and MTBE was obtained from the Journal of Applied Toxicology, and significant non-cancer health findings are reported below (47) (48). D-Limonene is a naturally derived citrus solvent and is often used as a cleaning agent, and MTBE is used to replace chemicals in chromatography and extraction processes (46).

*Table 22: MTBE short-term data.*

| Species | Sex | Endpoint | Dose (mg/kg) | N | Final BW | SD |
|---|---|---|---|---|---|---|
| Rats | male | BW | 0 | 10 | 434.6 | 42.4 |
| Rats | male | BW | 37 | 10 | 415.8 | 27.9 |
| Rats | male | BW | 209 | 10 | 422.7 | 36.9 |
| Rats | male | BW | 514 | 10 | 388.6 | 40.9 |
| Rats | male | BW | 972 | 10 | 385.6 | 39.1 |
| Rats | female | BW | 0 | 10 | 228.1 | 11.8 |
| Rats | female | BW | 50 | 10 | 239.2 | 18.5 |
| Rats | female | BW | 272 | 10 | 234.4 | 18.6 |
| Rats | female | BW | 650 | 10 | 229.7 | 20.4 |
| Rats | female | BW | 1153 | 10 | 240.8 | 20.9 |
| Rats | male | Kidney OW | 0 | 10 | 2.526 | 0.202 |
| Rats | male | Kidney OW | 37 | 10 | 2.592 | 0.217 |
| Rats | male | Kidney OW | 209 | 10 | 2.704 | 0.303 |
| Rats | male | Kidney OW | 514 | 10 | 2.718 | 0.138 |
| Rats | male | Kidney OW | 972 | 10 | 2.669 | 0.365 |
| Rats | female | Kidney OW | 0 | 10 | 1.416 | 0.106 |
| Rats | female | Kidney OW | 50 | 10 | 1.542 | 0.148 |
| Rats | female | Kidney OW | 272 | 10 | 1.554 | 0.127 |
| Rats | female | Kidney OW | 650 | 10 | 1.559 | 0.133 |
| Rats | female | Kidney OW | 1153 | 10 | 1.639 | 0.208 |
| Rats | female | Ovaries OW | 0 | 10 | 0.139 | 0.022 |
| Rats | female | Ovaries OW | 50 | 10 | 0.113 | 0.036 |
| Rats | female | Ovaries OW | 272 | 10 | 0.128 | 0.017 |
| Rats | female | Ovaries OW | 650 | 10 | 0.134 | 0.018 |
| Rats | female | Ovaries OW | 1153 | 10 | 0.14 | 0.024 |

BW= body weight
OW= organ weight
(47)

*Table 23: d-Limonene short-term data.*

| Species | Sex | Endpoint | dose (mg/kg) | N | final body weight | SD |
|---|---|---|---:|---:|---:|---:|
| Rats | male | BW | 0 | 10 | 333 | 18.97 |
| Rats | male | BW | 150 | 10 | 332 | 12.65 |
| Rats | male | BW | 300 | 10 | 330 | 9.49 |
| Rats | male | BW | 600 | 10 | 314 | 15.81 |
| Rats | male | BW | 1200 | 10 | 292 | 15.81 |
| Rats | male | BW | 2400 | 5 | 255 | 22.36 |
| Rats | female | BW | 0 | 10 | 185 | 6.32 |
| Rats | female | BW | 150 | 10 | 186 | 6.32 |
| Rats | female | BW | 300 | 10 | 181 | 6.32 |
| Rats | female | BW | 600 | 10 | 184 | 6.32 |
| Rats | female | BW | 1200 | 10 | 182 | 9.49 |
| Rats | female | BW | 2400 | 1 | 164 | 3.00 |
| mice | male | BW | 0 | 10 | 37.1 | 3.16 |
| mice | male | BW | 125 | 10 | 37.9 | 3.48 |
| mice | male | BW | 250 | 9 | 33.9 | 2.40 |
| mice | male | BW | 500 | 7 | 34.4 | 2.38 |
| mice | male | BW | 1000 | 9 | 33.3 | 2.40 |
| mice | male | BW | 2000 | 9 | 33 | 2.40 |
| mice | female | BW | 0 | 10 | 24.7 | 1.58 |
| mice | female | BW | 125 | 9 | 25.9 | 1.50 |
| mice | female | BW | 250 | 10 | 25.4 | 1.90 |
| mice | female | BW | 500 | 9 | 24.9 | 1.50 |
| mice | female | BW | 1000 | 10 | 24.1 | 2.21 |
| mice | female | BW | 2000 | 8 | 24.1 | 1.13 |

(48)

Next, STEP 2 of the methodology, the short-term toxicity data for d-Limonene and MTBE were modeled in BMDS. Both of these MIT chemical alternatives resulted in BMDS successfully modeling exactly one endpoint. Therefore both chemicals entered Transfer Point 1 and the single best-fit BMDL10 (short-term) per chemical alternative is the default minimum best-fit BMDL10 for that chemical. D-limonene's best-fit, minimum BMDL10 (short-term) was 138.64 mg/kg BW, and MTBE's is 502.90 mg/kg BW. These values were then converted to predicted chronic BMDL10s (51). See Table 24 below.

Now, as indicated in Figure 5 STEP 3, the availability of experimentally derived oral PODs for the targets to d-limonene and MTBE are assessed.  The only MIT proposed target for d-limonene, where a POD is easily accessible, is toluene.  Toluene is a solvent that is recognized as a developmental toxicant and air pollutant (46).  The POD was reported by the EPA's Integrated Risk Information System (IRIS), (see Table 24). Similarly, only one of the potential targets for MTBE has an easily obtained experimentally derived POD.  IRIS reports that the experimentally derived short-term BMDL10 for chloroform is 502.90 kg/mg BW. Chloroform is a volatile, hazardous organic solvent (46).  This value was converted into a predicted chronic value using the relationship's established by Kratchman, *et al*. (51).  While converting from a short-term POD to a chronic was not ideal, this was the only POD that could be obtained for any of the potential MTBE targets.

The final step in the methodology, represented by Transfer Point 2, is to compare the chronic BMDL10s of the target-alternative pairs to determine which chemical is predicted to be more hazardous to human health, in terms of non-cancer health endpoints. The chemical in the pair with the higher BMDL10, can be considered "safer", within the context of this study.

*Table 24: Results of Transfer Points 1 and 2; comparison of target-alternative pairs.*

| Transfer Point 1:<br>Target- alternative BMDL10 | | | | Transfer Point 2:<br>target-alternative comparison |
|---|---|---|---|---|
| Target | Chronic BMDL10 (mg/kg BW) | Alternative | Predicted chronic BMDL10 (mg/kg BW) (51) | Highest BMDL10 (i.e. “safer” chemical of pair) |
| Toluene | BMDL10$^{=}$ 164 | d-Limonene | 27.86* | **Target (toluene)** |
| Chloroform | BMDL10= 1.0 | MTBE | 253.18 ** | **Alternative (MTBE)** |

* Male mice BW; best-fit model: exponential, 4th power
** Male rat BW; best-fit model: exponential, 2nd power

## 5 Discussion

This study demonstrates that in the absence of chronic non-cancer reference values and bioassays, there is a need for predictive techniques and consistent methodologies for considering alternative-target chemical pairs with data imbalances (13). Based on the results presented in Table 24, replacing toluene (target) with d-limonene (alternative) is not a safer alternative (“safer” in terms of human toxicity); toluene is predicted to have a higher BMDL10 than d-limonene. On the other hand, MTBE (alternative) is predicted to be safer than chloroform (target), because the BMDL10 for MTBE is higher than that of chloroform. This variability, even within the same framework, underscores the need for consistent methodologies (4) for assessing chemical alternatives, and in the absence of chronic toxicity data, the need for quantitative techniques in predicting chronic non-cancer toxicity levels as Kratchman, *et al*. (Chapter 2) suggest (51).

A surprising result of this study is that out of 88 possible target-alternative pairs only two could be identified where the alternative had short-term data and could be modeled in BMDS, and the target had experimentally derived chronic toxicity data. This was not an expected result and demonstrates that in addition to the need for more chronic toxicity data or methods to predict chronic effects (51) there is also a deficiency of *short-term* toxicity data. It again should be emphasized again that the MIT framework was assessed because it is prototypical of chemical alternative tools available to the public. It

There is uncertainty associated with the predictions from the Kratchman, *et al*. relationships (Chapter 2), however the differences between BMDL10 estimates for the target and alternative pairs was quite large. This provides some confidence in the comparisons. Further work could develop quantitative estimates of uncertainty in BMDL10 estimates for both target and alternative chemicals. Another area of future research could be in comparing actual reference doses used in practice, to references doses that could be calculated by applying UFs to the BMDs and BMDLs predicted using Kratchman *et al.*'s relationships (Chapter 2, Equation 1). The major challenge however, is the amount of uncertainty that would accrue following the use of the predictive relationships as well as UFs, all of which would result in large amounts of error.

With such data gaps and inconsistencies in findings within one prototypical framework, it is necessary to continue to pursue the development of chronic toxicity data and reference values, and in the absence of such data, to develop quantitative techniques that predict chronic health outcomes when data disparities exist. (51) (13) (4).

## 5.1 Study limitations

One limitation of this study is assuming that the differences in relative exposure between target and alternative chemicals are negligible. This is because the citations used by MIT often were not based on quantitative Further, the assumption that higher BMDL10 is safer (from a human toxicity perspective) can be a limitation if both chemicals within the pair are far enough below a threshold that both are actually safe.

With a marked lack of data available for both chronic and short-term only two chemical pairs were analyzed. This presents a limitation with the predictive methodology presented by Kratchman, *et al*., and with BMDS software. This also further highlights the need for more chronic health references values and toxicity data.

Another limitation is that the correlations developed by Kratchman, *et al*. apply only to non-cancer endpoints. This could be a limitation in determining regrettable substitutions, especially in instances where cancerous endpoints were the most critical effect (2). Further, the equations presented by Kratchman (Chapter 2) are based upon endpoints identified by NTP pathologists as either biologically or statistically significant; however in the case of MTBE, this paper did not use NTP data. That could introduce comparability issues in how the NTP versus the Journal of Applied Toxicology pathologists identified "significant" endpoints. It must also be noted that Kratchman, *et al.* equations are based upon 41 chemicals. It was not assess for this study if that was sufficient to extrapolate a relationship that can be used to predict any random chemicals.

## 5.2 Conclusions

The absence of chronic non-cancer reference values and bioassays is a major hurdle in developing effective and consistent chemical alternatives assessments which identify "safer" alternatives. This study considers two target-alternative pairs presented by MIT, considered a prototypical alternatives assessment framework. Safer within this study was defined as the chemical with the highest BMDL10 amongst a target-alternative chemical pair. While one of the pairs assessed met this criterion (MITBE and chloroform), the other pair did not (d-limonene and toluene). This demonstrates inconsistencies in methodologies of comparing chemicals. It also highlights the need for methods to rapidly estimate chronic health reference values in a consistent and reliable manner.

# Chapter 5. Conclusions

The main goal this dissertation research was to develop and assess quantitative relationships for predicting doses associated with chronic non-cancer health outcomes in situations where there is an absence of chronic toxicity data, and to consider the application of these findings to chemical substitution decisions. The effectiveness of short-term toxicity data as a predictor of chronic toxicity effect levels was considered with a focus on dose-response relationships instead of a critical health effect. Data from NTP TRs were extracted and modeled using EPA's BMDS. Best-fit, minimum (and median) BMDs and BMDLs were determined for all NTP pathologist identified significant non-neoplastic lesions, final mean body weights and mean organ weights for 41 chemicals tested by NTP between 2000 and 2012.

The three studies carried out in this dissertation research each addressed a different step associated with characterizing non-cancer human toxicity levels from chemicals. The relationships assessed can only be applied to non-cancer endpoints. Including cancerous endpoints in the comparisons could be the subject of future research. Paper 1 assessed the relationship between short-term and chronic health data and the ability to predict the doses associated with chronic health outcomes from short-term data. Paper 2 considered the next phase of chemical assessment by considering sensitivity of species and species-sex groups. Finally, the findings of Paper 1 (Chapter 2) were considered on a practical level in Paper 3 (Chapter 4), where target-alternative chemical pairs recommended by a prototypical chemical substitution framework were assessed.

The findings of Paper 1 (Chapter 2) indicate that regardless of the chemical or health endpoint, short-term animal studies reasonably provide a quantitative estimate of chronic non-cancer toxicity levels including minimum and median values of BMD10, BMDL10, BMD50 and BMDL50. This can allow for faster development of PODs for use in developing reference values like the reference doses or margins of exposure (21). The findings are robust, and true for continuous, dichotomous and absolute datasets, as well as between minimum or median BMD and BMDL findings.

In Paper 2 (Chapter 3), it was determined that for the group of chemicals assessed there is not a uniform species and species-sex sensitivity distribution. This was true across chemicals and is not dependent on endpoint type. The implications of these findings could affect future toxicity testing guidelines and encourage more rigorous consideration of the selection of species-sex test subjects in general toxicity testing guidelines.

The final study, Paper 3 (Chapter 4), considered two target-alternative pairs presented by a prototypical chemical alternatives framework. While one of the pairs assessed were determined to be a safe substitution in terms of human health toxicity (MITBE and chloroform), the other pair appears not to be (d-limonene and toluene). This demonstrates inconsistencies in methodologies of comparing relative toxicity levels in data sparse situations. It also highlights the need for methods to rapidly estimate chronic health reference values in a consistent and reliable manner.

The findings of all three of these papers indicate that while the desire for finding *safer* chemicals is increasing, there are limitations with the current state-of-knowledge within each of the major steps of the chemical assessment process. At the data acquisition step, as Paper 1 (Chapter 2) demonstrated, chronic non-cancer toxicity data is not always available, leading to the need for more research of predictive techniques that are consistent and empirical. The next step, assessment of sensitivities, was investigated by Paper 2 (Chapter 3). The findings indicate that chemical assessors for the group of chemicals in this study may not have a uniform distribution of species-sex sensitivity. Finally, at the application step, represented by Paper 3 (Chapter 4), there are inconsistencies within chemical substitution frameworks. The lack of data and assessment of relative sensitivities can lead to subjectivity and inconsistencies not only between chemical substitution frameworks but within them as well.

Combined, the three papers in this dissertation have broad applicability. The findings and relationships developed can be used to facilitate rapid single chemical assessments, and assess the human health impacts data sparse chemicals. In addition to single chemical assessments, these findings can facilitate the assessment of relative toxicity between chemicals such as alternatives assessments.

# References

1. *Prediction of Rodent Carcinogenesis: An Evaluation of Prechronic Liver Lesions as Forecasters of Liver Tumors in NTP Carcinogenicity Studies.* Allen, D.G., et al. p. 393-401, s.l. : Toxicologic Pathology, 2004, Vol. 32.

2. *Concordance of Noncarcinogenic Endpoints in Rodent Chemical Bioassays.* Wang, B. and Gray, G. 2015, Risk Analysis, Vol. 35, No. 6, pp. 1154-1166.

3. U.S. Environmental Protection Agency. *Chemical Alternative Assessment.* [Online] 11 20, 2014. [Cited: 06 24, 2014.] http://www.epa.gov/sustainability/analytics/chem-alt.htm.

4. *Alternatives Assessment Frameworks: Research Needs for the Informed Substitution of Hazardous Chemicals.* Jacobs, M., et al. 3, s.l. : Journal of Environmental Health Perspectives, 2016, Vol. 124.

5. *Alternatives Assessment: The Science of Identifying Safer Substitutes: Webinar for Change.* University of California, Los Angeles, Law and Environmental Health. Los Angeles : s.n., 2009.

6. U.S. Environmental Protection Agency. *Risk Assessment for Noncancer Effects.* [Online] 02 05, 2016. [Cited: 09 19, 2016.] https://www.epa.gov/fera/risk-assessment-noncancer-effects.

7. *Assessment of the Oral Rat Chronic Lowest Observed Adverse Effect Level Model in a QSAR Software Package for Toxicity Prediction.* Venkatapathy, R., Moudgal, C. J. and Bruce, R. M. 5, s.l. : Journal of Chemical Information Computer Science, 2004, Vol. 44.

8. U.S. Environmental Protection Agency. IRIS Advanced Search. [Online] 03 2017. https://cfpub.epa.gov/ncea/iris/search/.

9. —. Vocabulary Catelog: Integrated Risk Information System (IRIS) Glossary. *U.S. Environmental Protection Agency.* [Online] October 05, 2016. https://iaspub.epa.gov/sor_internet/registry/termreg/searchandretrieve/glossariesandkeywordlists/search.do;jsessionid=W6f8l-Imgy2jhzQyyDqWbomSlZEihgkQNMZNd8r9rbbMf86pgLEs!-1900823099?details=&vocabName=IRIS%20Glossary&filterTerm=point%20of%20departure&che.

10. *Guide to the Care and Use of Laboratory Animals.* Canadian Council on Animal Care. 1993, Vol 1, p. 122.

11. OECD. *Organization of Economic Cooperation and Development: Guidelines for the Testing of Chemicals: Test Guideline 452: Chronic Toxicity Studies.* Paris, Fr : OECD, 2008.

12. *An Overview of Prechronic and Chronic Toxicity/Carcinogenicity Experimental Study Designs and Criteria Used by the National Toxicology Program.* Chhabra, R.S., et al. 1990, Environmental Health Perspectives, pp. Vol. 86, 313-320.

13. *Setting Pesticide Reference Doses: A Retrospective Analysis Examining Key Data and Choices.* . Holman, E., Francis, R. and Gray, G. s.l. : Human and Ecological Risk Assessment, 2016, Vols. 20:1550-1564.

14. *Transforming Environmental Health Protection.* Collins, Francis S., Gray, George M. and Bucher, John R. 5865, s.l. : Science, 2008, Vol. 319.

15. Tox21 predicted to have dramatic impact of US EDSP test orders. *News, Chemical Watch: Global Risk and Regulation.* 2013.

16. *Assessing Human Health Response in Life Cycle Assessment Using ED10s and DALYs: Part 2—Noncancer Effects.* Pennington, David, et al. 5, s.l. : Journal of Risk Analysis, 2002, Vol. 22.

17. Lewis, David F.V. Quantitative Structure Activity Relationship. *Computer-Assisted methods in the evaluation of chemical toxicity.* s.l. : Reviews in Computation Chemistry, 2007, Vol. 3.

18. National Toxicology Program. National Toxicology Program. *NTP Reports and Publications.* [Online] June 26, 2016. http://ntp.niehs.nih.gov/results/pubs/index.html.

19. U.S. Department of Health and Human Services. National Toxicology Program. *Listing of Short-Term Toxicity Reports.* [Online] 08 25, 2016. [Cited: 11 21, 2016.] https://ntp.niehs.nih.gov/results/pubs/shortterm/reports/index.html.

20. U.S. Environmental Protection Agency. *BMDS 2.6.0.* Washington, D.C. : U.S. Environmental Protection Agency, 2015.

21. U.S. Environmental Protection Agency: Risk Assessment Forum. *Benchmark Dose Technical Guidance.* Washington, D.C. : U.S. EPA, 2012. EPA/100/R-12/001.

22. *Current modeling practice may lead to falsely high benchmark dose estimates.* Ringblom, J., Johanson, G. and M.Öberg. 2, s.l. : Journal of Regulatory Pharmocology and Toxicology, 2014, Vol. 69.

23. *The impact of model uncertainty on benchmark dose estimation. .* West, R. Webster. s.l. : Environmetrics, , 2012, Vol. 23.

24. *A procedure for developing risk-based reference doses.* Gaylor, D. and Kodell, R. p. 137-141, s.l. : Journal of Regulatory Toxicology, 2002, Vol. 35.

25. Technology, National Institute of Standards and. Engineering Statistics Handbook. *Quantile-Quantile Plot.* Gaithersburg, MD : NIST, 2013.

26. *Prediction intervals in linear regression taking into account errors on both axes. .* Rio, F.J. del, Riu, J. and Rius, F.X. 773-788 , s.l. : J. of Chemometrics , 2001, Vol. 15.

27. Minitab17 Support. *Orthogonal Regression.* [Online] Minitab Inc., 2016. [Cited: 08 16, 2016.] http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/regression-and-correlation/basics/orthogonal-regression/.

28. Casella, G. and Berger, R. *Statistical Inference.* Pasific Grove, CA : Brooks-Cole Publishing Company, 1990.

29. *The Use and Misuse of Orthogonal Regression in Linear Errors-in-Variables Models.* Carroll, R. and Ruppert, D. 1, s.l. : The American Statistician, 1996, Vol. 50.

30. *Lines, models, and errors; regression in the field.* McArdle, BH. 1363–1366., s.l. : Limnology and Oceanography, 2003, Vol. 8.

31. *Use and misuse of the reduced major axis for line-fitting.* . Smith, R.J. 476-486, s.l. : American Journal of Physical Anthropolology, 2009, Vol. 140.
32. R Core Team. R: A language and environment for statistical computing. [Online] The R Foundation for Statistical Computing, 2016. https://www.r-project.org/.
33. GraphPad Software. The distinction between confidence intervals, prediction intervals and tolerance intervals. *GraphPad Software Support.* [Online] 2017. [Cited: 03 26, 2017.] https://www.graphpad.com/support/faqid/1506/.
34. National Toxicology Program. Chemical Effects in Biological Systems Database. *Search Study.* [Online] 01 2016. [Cited: 01 15, 2016.] http://tools.niehs.nih.gov/cebs3/ui/.
35. *Use of BMDS and PROAST software packages by EFSA Scientific Panels and Units for applying the Benchmark Dose (BMD) approach in risk assessment.* European Food Safety Authority. 2011, pp. EN-113.
36. Gad, S.C. *Animal Models in Toxicology, Second Edition.* Boca Raton, FL : CRC Press, 2006.
37. *Rats!* Iannaccone, P. and Jacob, H. 2009, Disease Models and Mechanisms, pp. 206-210.
38. Canadians for Health Research. Rats and Research. *Canadians for Health Research.* [Online] June 28, 2016. http://www.chrcrm.org/en/rats-and-research.
39. *Thalidomide: The Tragedy of Birth Defects and the Effective Treatment of Disease.* Kim, J. and Scialli, A. 2011, Toxicological Sciences, pp. 1-6.
40. U.S. Department of Health and Human Services. National Toxicology Program. *Listing of Short-Term Toxicity Reports.* [Online] 2016. [Cited: 05 15, 2016.] http://ntp.niehs.nih.gov/results/pubs/shortterm/reports/index.html.
41. National Toxicology Program. *NTP Technical Report on the Toxicology and Carcinogenesis Studies of Elmiron® (CAS No. 37319-17-8) in F344/N Rats and B6C3F1 Mice (Gavage Studies).* s.l. : NIH, 2004. 04-4446.
42. *Decisions, Science, and Values: Crafting Regulatory Alternatives Analysis.* T. Malloy, A. Blake, I. Linkov and Sinsheimer, P. 12, s.l. : Journal of Risk Analysis, 2015, Risk Analysis, Vol. 35, pp. Vol. 35, 12, p.2137-2151.
43. *A Framework to Guide Selection of Chemical Alternatives.* National Research Counsil. Washington, D.C. : The National Academies Press, 2014.
44. Occupational Safety and Health Administration. *Transitioning to Safer Chemicals: A Toolkit for Employers and Workers.* [Online] U.S. Department of Labor. https://www.osha.gov/dsg/safer_chemicals/why_transition.html.
45. *Policy: Rethink chemical risk assessments.* Gray, G. and Cohen, J. s.l. : Nature, 2012, Vol. 489. 10.1038/489027a .
46. "Green" Alternatives Wizard. [Online] Massachusetts Institute of Technology, Department of Environmental Health. http://ehs.mit.edu/greenchem/.

47. *Toxicity of methyl tertiary-butyl ether (MTBE) following exposure of Wistar Rats for 13 weeks or one year via drinking water.* Bermudez, E., et al. s.l. : Journal of Applied Toxicology, 2011, Vol. 32.

48. National Toxicology Program. *Toxicology and Carcinogensis Studies of d-Limonene in F344/N Rats and B6C3F1 Mice.* Washington, D.C. : National Institutes of Health, 1990. NTP347.

49. Integrated Risk INformation System (IRIS). *Chemical Assessment Summary: Chloroform; CASRN 67-77-3.* Washington, D.C. : U.S. Environmental Protection Agency, 1987.

50. Integrated Risk Information System (IRIS). *Chemical Assessment Summary: Toluene; CASRN 108-88-3.* Washington, D.C. : U.S. Environmental Protection Agency, 1990.

51. *Correlation of Non-Cancer Benchmark Doses in Short and Long-Term Rodent Bioassays.* Kratchman, J., et al. s.l. : Journal of Risk Analysis, 2017 [under review].

52. U.S. Environmental Protection Agency. *IRIS Advanced Search.* [Online] 12 29, 2016. https://cfpub.epa.gov/ncea/iris/search/.

53. American Chemcial Society. Common Chemistry: A CAS Solution. [Online] 12 2016. http://www.commonchemistry.org/index.aspx.

54. Google Inc. Google. *Google Scholar.* [Online] http://scholar.google.com/.

55. *Incorporating New Technologies into Toxicity Testing and Risk Assessment: Moving from 21st Century Vision to a Data-Driven Framework.* Thomas, Russell S., et al. 1, s.l. : Toxicological Sciences, 2013, Vol. 136.

56. *Methods for Identifying a Default Cross-Species Scaling Factor.* Rhomberg, L. R. and Lewandowski, T. A. 2004, U.S. Environmental Protection Agency Risk Assessment Forum.

57. *Animal Models of Toxicity: Some Comparative Data on the Sensitivity of Behavioral Tests.* Reiter, L., et al. Ohio : s.n., 1980. Proceedings of the Eleventh Conference on Environmental Toxicology 18, 19, and 20 November 1980. p. 11.

58. Lewis, M. "Alternatives Assessments Key to Manufacturers as California Green Chemistry Initiative Kicks in". *ThomasNet News.* 2013.

59. *A straw man proposal for a quantitative definition of the RfD.* Hattis, D., Baird, S. and Goble, R. http://www.ncbi.nlm.nih.gov/pubmed/12378950,

60. U.S. Environmental Protection Agency. *Toxicity Testing in the 21st Century (Tox21).* [Online] 08 17, 2016. [Cited: 08 30, 2016.] https://www.epa.gov/chemical-research/toxicology-testing-21st-century-tox21.

61. —. EPA Connect: Design for the Environment Teams up with Business Leaders. [Online] 06 12, 12 May 2014. [Cited: 06 25, 2014.] http://blog.epa.gov/epaconnect/2014/05/design-for-the-environment-teams-up-with-business-leaders/.

62. ABC News San Francisco. Debate Underway Over California's Green Chemistry Revolution. [Online] ABC News, 08 12, 2009. [Cited: 08 01, 2015.] http://abc7news.com/archive/7160331.

63. Stat Trek. Chi-Squared Goodness of Fit Test. [Online] May 18, 2016. http://stattrek.com/chi-square-test/goodness-of-fit.aspx?Tutorial=AP.

64. Walmart. "Walmart Highlights Progress on the Sustainability Index,". [Online] 12 09, 2013. [Cited: 08 31, 2014.] http://news.walmart.com/news-archive/2013/09/12/walmart-highlights-progress-on-the-sustainability-index.