Mixed Effects Modeling Methods to Find ten Berge Exponents

Lizzy Compton, Jacob Smith

November 2, 2020

Executive Summary

As outlined in the Standing Operating Procedure for the National Institute for Occupational Safety and Health (NIOSH), linear regression for the relationship between log time and log LC_{50} values can be used to estimate the chemically-dependent ten Berge exponents. In an effort to establish Immediately Dangerous to Life and Health (IDLH) values for chemicals deemed dangerous in the workplace, data from decades of experiments have been collected and analyzed. While ensuring to account for Species, the simple linear regression model was expanded to a mixed effects model that takes study to study variability into account. Exponent estimates and bootstrapped confidence intervals for chemicals with enough data points were computed. Carbon disulfide and Epichlorohydrin have exponents around 0.85; Pentaborane and Ammonia have exponents around 1.3; and Carbon monoxide and Hydrogen cyanide have exponents around 1.7. These values were compared to estimates using the standard linear regression, yielding no apparent improvement in model selection. Results are discussed more in depth along with the main hurdle in this study being the extremely small sample sizes for many of the chemicals.

Introduction

The National Institute for Occupational Safety and Health (NIOSH) is a branch within the Center for Disease Control and Prevention (CDC) that uses toxicology to evaluate the risk of various chemicals. When dealing with dangerous chemicals, it is important to know how the concentration and exposure time relate together to understand the toxicity of the chemical. When chemicals are inhaled reactions such as irritation, impaired breathing and/or vision, serious health effects and even death can occur. As a precaution, NIOSH sets IDLH (Immediately Dangerous to Life and Health) values. These values indicate the concentration limit for a chemical that allows for the proper time needed for someone to evacuate if the work environment becomes contaminated by the chemical (cdc.gov).

In order to set the IDLH values, NIOSH must know the relationship between a chemical's concentration, the time exposed to the chemical and the toxicity. Historically, Haber's Law established that toxicity is the direct product of concentration and exposure time. This assumed that this relationship is consistent for all chemicals, that damage is irreversible, and toxicity is cumulative. This was considered precedent until (ten Berge et al., 1986*)* argued that this relationship will change from chemical to chemical. To account for this, ten Berge suggested first raising concentration to a chemically dependent exponent and then multiplying it by the exposure time to find the toxicity. Next, linear regression is used to model the log log transformation of concentration and time and this model is then used to solve for the exponent (referred to as the ten Berge n exponent) for each individual chemical. This method is now considered the Standing Operating Procedure (National Research Council, 2001).

Both methods here assume that concentration and exposure time are the only two components that we need to account for when calculating the toxicity. However, the data used

within these calculations often comes from multiple studies and this is an additional piece of variation that needs to be accounted for. Thus, this report will use the ten Berge method but will include the added component of adjusting for the variation between each study. We will create mixed effects models for each of the chemicals provided and calculate an estimate for the ten Berge exponent for each. Following this, we will use bootstrapping methods to create confidence intervals for each of the estimates of n. Lastly, we will compare the estimates and intervals created from the mixed effects models to those of the traditional ten Berge method to determine the effect that accounting for the reference study has.

Data

NIOSH provides data tables of a whole host of chemicals on the CDC website. These data tables give the lethal concentration data from prior studies that have been conducted and recorded. The data provides the species the data was collected on, the concentration (given as an LC_{50} value), the exposure time and the reference study in which the data was collected. Species are typically small mammals, but our report will focus on the Rat species as those typically have the largest sample sizes. LC_{50}indicates the lethal concentration in which 50% of the population dies. These values are measured in parts per million (ppm) or milligrams per cubic meter (mg/m^3). The exposure time values range from 1 minute to 8 hours but for our analysis purposes, all time values have been converted to hours (cdc.gov).

We have been provided with thirty chemicals pulled from the database available on the CDC website. Of those thirty chemicals, only eight had at least three observations with unique LC_{50} values and unique time values. This is a requirement to conduct the linear regression analysis. These eight chemicals are listed below.

Table 1: List of Chemicals and Observations

Chemical Name	Species	Unique LC_{50} values	Time values	Reference Studies
Ammonia	Rat	6	5	3
Acrolein	Rat	3	3	3
Carbon disulfide	Rat	4	3	2
Carbon monoxide	Rat	3	3	2
Epichlorohydrin	Rat	5	3	4
Ethyleneimine	Rat	3	3	3
Hydrogen cyanide	Rat	5	3	4
Pentaborane	Rat	5	5	2
Pentaborane	Mouse	6	5	3

However, since Acrolein and Ethyleneimine both only have three unique observations, with each coming from its own reference study, neither can be analyzed using a mixed model. In order to account for the variation from the studies, there needs to be at the very least two observations from at least one study. Thus, the final chemicals used for the mixed modeling methods are Ammonia, Carbon disulfide, Carbon monoxide, Epichlorohydrin, Hydrogen cyanide and Pentaborane. Since Pentaborane has multiple observations for the rat and mouse species, we create a ten Berge estimate for both and compare across.

Methods

In 1924 Haber's Law established that the toxicity of a chemical is the direct product of the concentration and exposure time.

$$C * t = k$$

This assumes that the damage is irreversible, and that toxicity is cumulative. Additionally, it was assumed that this relationship remained constant regardless of the chemical. (Dotson et al.,

2012). However, in 1986 ten Berge et al. argued that there are chemically specific relationships between the exposure concentration and the duration of that exposure. They proposed the following relationship,

$$C^n * t = k$$

where C is the concentration, t is the exposure time and k is the toxicity constant. The n exponent, known as the ten Berge exponent, is chemical-dependent and can be determined from a linear regression analysis of the log log transformation of concentration versus time. Starting with this equation, we can take the log of both sides and rearrange the equation to get:

$$C^n * t = k$$

$$\log(C^n * t) = \log(k)$$

$$n * \log(C) + \log(t) = \log(k)$$

$$n * \log(C) = \log(k) - \log(t)$$

$$\log(C) = \frac{\log(k)}{n} - \frac{1}{n}\log(t)$$

When creating a linear regression analysis of the log log transformation of concentration versus time we have a model of the form

$$\log(C) = \beta_0 + \beta_1 \log(t) + \varepsilon$$

Thus, by assuming the ten Berge relationship is modeled by this linear regression analysis, we can estimate n by setting $\beta_1 = -\frac{1}{n}$ (National Research Council, 2001).

In this report we will investigate the effect that the Reference study has on the relationships here. Oftentimes, since these studies are investigating the deadly effects of chemicals on live animals, very few data points are actually recorded. Thus, we must use data from multiple studies, and as a result, we then need to account for any variation between these

studies. Therefore, we propose the use of a mixed model with log(t) as the fixed effect, Reference study as a random effect and the log(C) as the response. This model has the form

$$log(C) = \beta_0 + \beta_1 log(t) + Study + \varepsilon$$

where Study ~ Normal(0, $\sigma^2{}_{Study}$) and ε ~Normal(0, $\sigma^2{}_{Error}$), and both are random effects. Additionally, we will assume that the Reference studies are independent random variables; that is, that one study is independent of all other studies. Since study is a random effect, it will be included as an additional variance component in the model. Therefore, the ten Berge relationship will still hold up and we can estimate n by setting $\beta_1 = -\frac{1}{n}$.

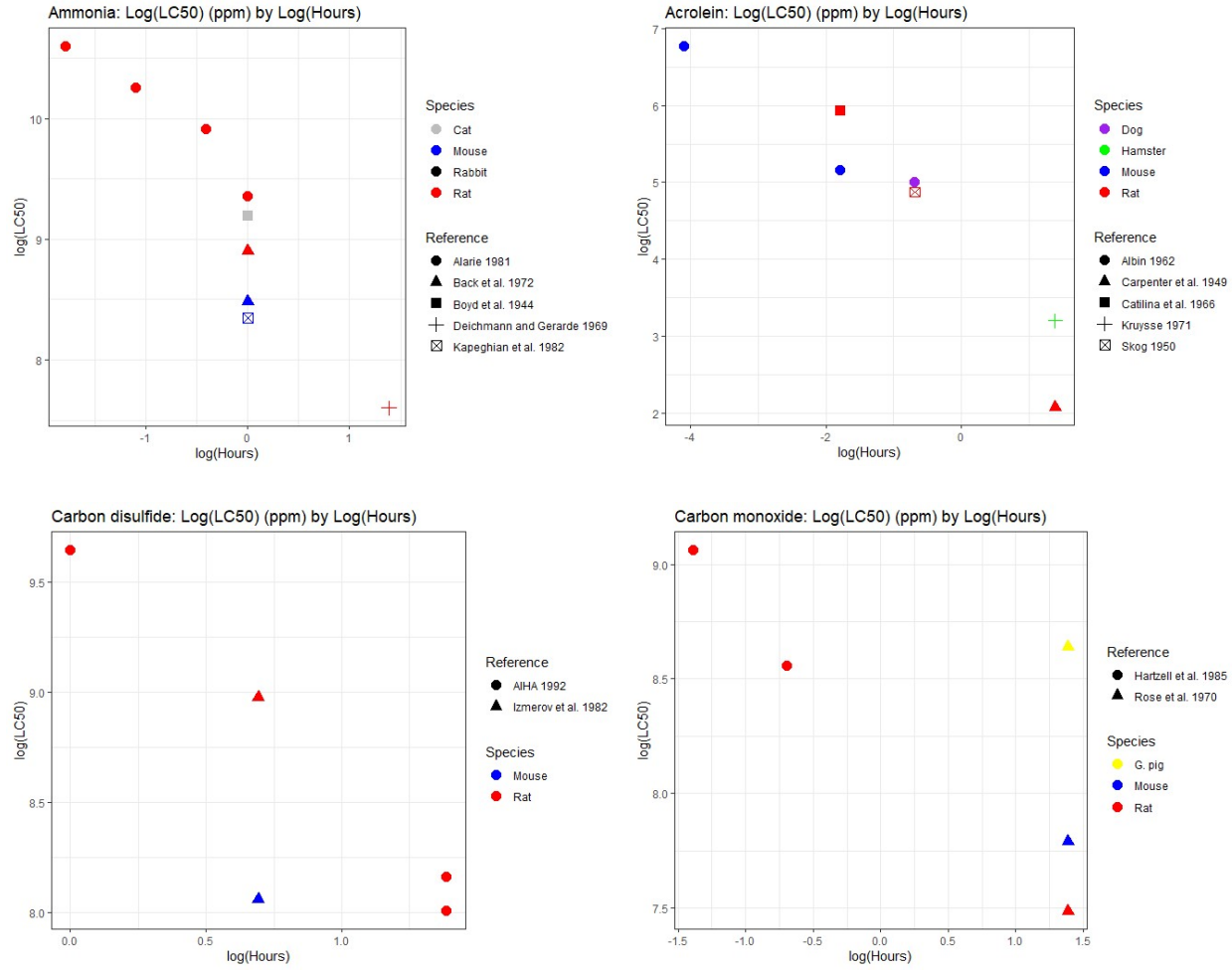
Following the calculation of the ten Berge, n, exponent estimates, we will also create bootstrap confidence intervals for these estimates. This will give us a range of plausible values for the estimate, n, instead of a single point estimate. Bootstrapping is a statistical process in which we randomly sample, with replacement, from an original sample to create multiple "bootstrap" samples. We can then use these samples to create a bootstrap distribution, from which we can calculate confidence intervals. Bootstrapping is often used when we do not meet the assumptions necessary to create traditional confidence intervals. Since we have such small sample sizes, this is the case here. We will use the lme4 package within R and the bootMer function to create the confidence intervals for n (The Comprehensive R Archive Network, 2020*)*.

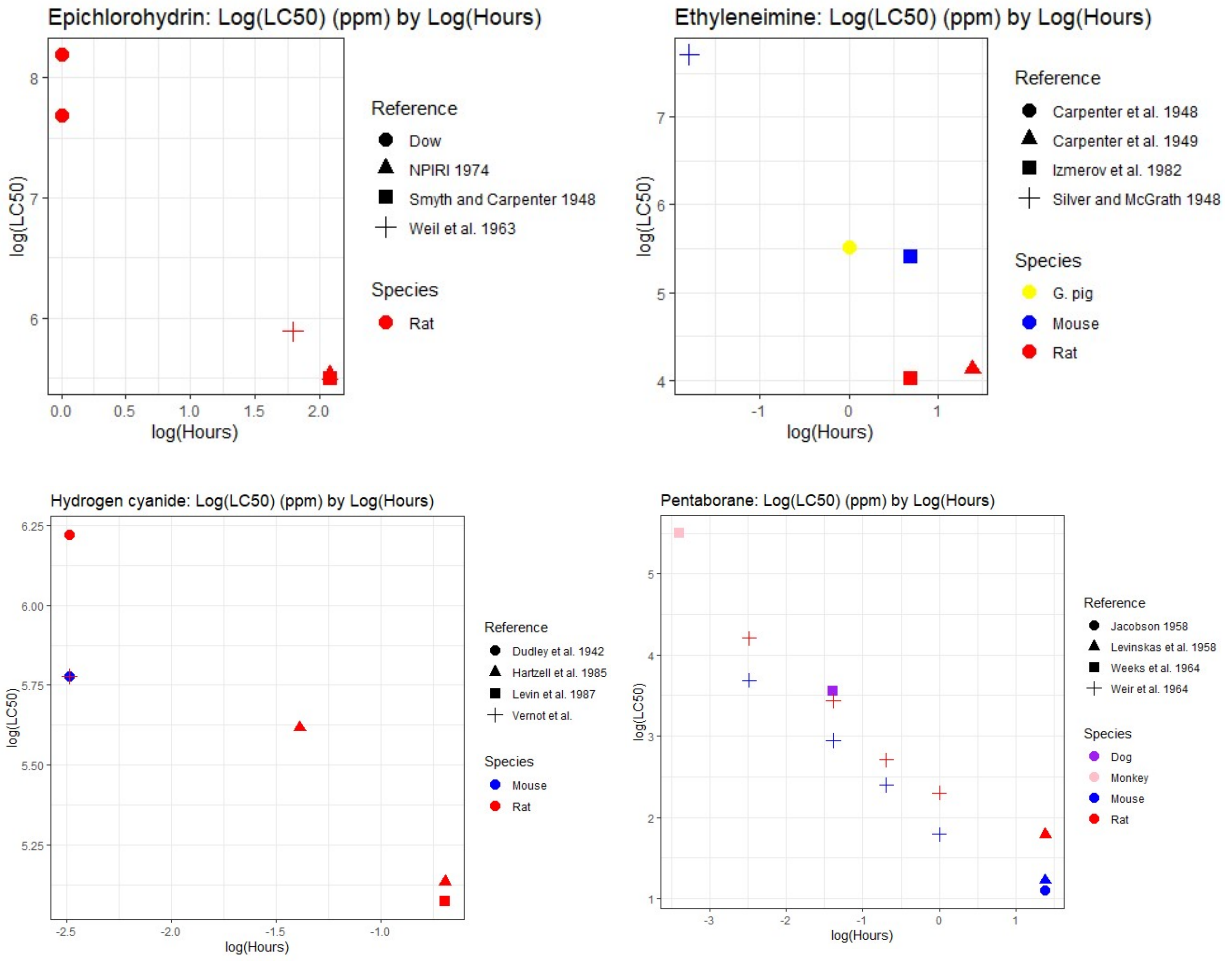
Exploratory Data Analysis

As mentioned in the Standing Operating Procedure, the relationship to be modeled is log LC_{50} by log time. The aforementioned eight chemicals (Table 1) of interest in this study have this relationship plotted below in Figure 1. Discussion on each individual plot will not be covered for the sake of repetition. However, important features to note from the plots as a whole are as follows. First, there is a clear downward trend in log LC_{50} values as log hours increases.

Second, note that the slope of the downward trend is different for each chemical and therein lies the chemically dependent ten Berge exponent estimate. And lastly, Reference studies are represented by shape for use as a random effect in the mixed effects model, and Species is represented by color allowing for easy recognition of the available sample sizes for each chemical. From these plots Ammonia and Pentaborane seem to the golden standard in terms of sample size and viable data points of the chemicals provided to us.

Figure 1: Log(LC_{50}) vs. Log(Time) for all chemicals





Results

Table 2 gives the resulting ten Berge exponent estimates for the six chemicals as well as the 95% bootstrapped confidence intervals. Pentaborane had adequate sample sizes for both the Rat and Mouse Species, so these estimates are given separately, otherwise all chemicals were filtered for the Rat Species.

Table 2: ten Berge Exponent Estimates

Chemical Name	Species	Number of Observations	Number of Time values	Estimate for n	Bootstrapped 95% CI Lower Bound	Bootstrapped 95% CI Upper Bound
Ammonia	Rat	6	5	1.397	1.060	1.988
Carbon disulfide	Rat	4	3	0.875	0.775	1.057
Carbon monoxide	Rat	3	3	1.775	1.095	5.342
Epichlorohydrin	Rat	5	3	0.860	0.743	1.029
Hydrogen cyanide	Rat	5	3	1.736	1.079	4.392
Pentaborane	Rat	5	5	1.288	1.144	1.467
Pentaborane	Mouse	6	5	1.341	1.221	1.500

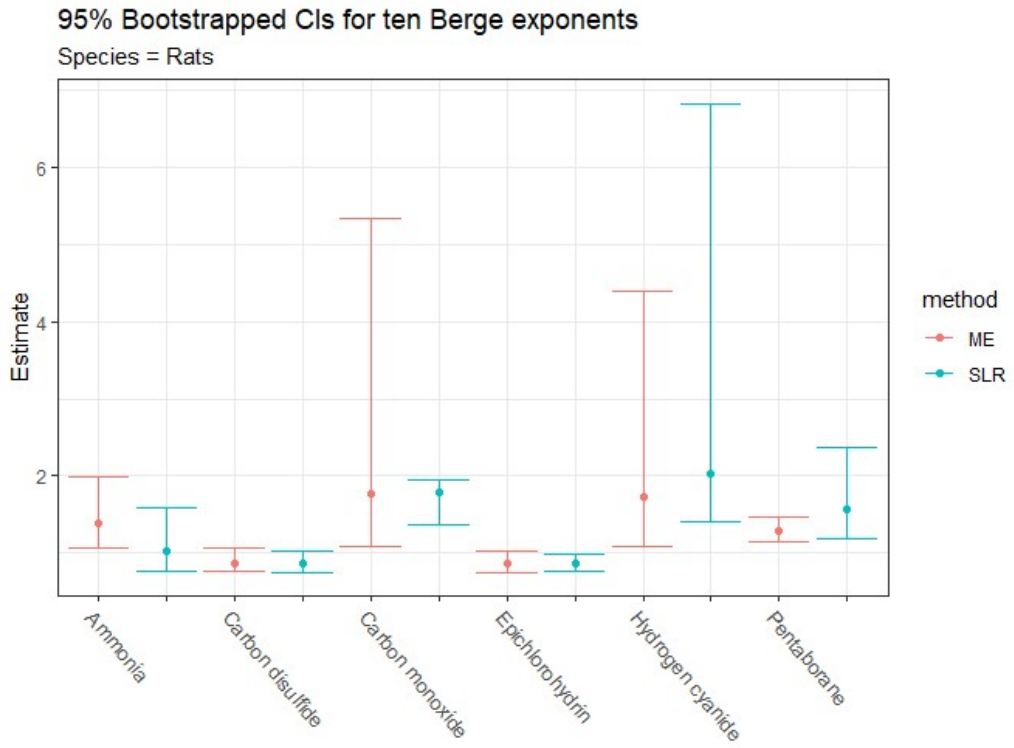
The estimates range in value from 0.86 to 1.775 and confidence interval widths range from about 0.7 up to 5.4. The 95% bootstrapped confidence interval for Ammonia can be interpreted as, we are 95% confident that the true ten Berge exponent for Ammonia is between 1.060 and 1.988. This interpretation can be generalized for the rest of the chemicals. According to the Standing Operating Procedures, typical ten Berge exponents fall within the range of 0.85 to 3.5. We should note that Carbon disulfide and Epichlorohydrin both have point estimates that fall on the edge of this range and their confidence intervals both exceed the range in the lower limit. Also, Carbon monoxide and Hydrogen cyanide both have upper limits that exceed this range, but their point estimates are well within the typical values. Thus, we should be aware of these four chemicals, but overall, there is little evidence to show that any of these exponents are outside the typical range of 0.85 to 3.5. Lastly, it is important to observe that Pentaborane produced similar exponent estimates between the Rat and Mouse Species. This indicates that the ten Berge

exponents are consistent within this one chemical, regardless of the Species. However, this does not provide strong evidence that this will be the case for all chemicals.

Discussion

A similar analysis was performed by Matthew Snyder and Takumi Kijima but without accounting for the variability from experiments in different References. This is essentially a simple linear regression model and the traditional way of calculating the ten Berge exponents according to the Standing Operating Procedures. To assess whether accounting for the variability of the Reference makes much of a difference in the estimates, the confidence intervals from each method, for the six chemicals are compared below for the Rat Species. For each chemical, the mixed effects method (shown in red) is next to the simple linear regression method (shown in blue) for comparison.

Figure 2: Model Comparisons



All the point estimates, indicated by the dot within the intervals, are pretty close or visibly the same as the estimates from the opposing method. Ammonia, Hydrogen cyanide, and Pentaborane, however, produce slightly different estimates between the two methods. Of main interest though is the change in variability between the methods. Ammonia, Carbon disulfide, and Epichlorohydrin all have similar sized confidence intervals between the two. On the other hand, there is an increased confidence interval width in the SLR method for Hydrogen cyanide and Pentaborane and vice versa for Carbon monoxide.

Through this analysis, we can see that there are advantages and disadvantages to both the mixed effects modeling and the simple linear regression methods. One advantage of the mixed effects model is that it better accounts for variation between studies and any correlation between data points within the same study. However, it requires more data and at least one study with more than one observation. Simple linear regression on the other hand is a simpler model to interpret and is easier to implement. For example, the SLR method is able to compute estimates for Acrolein and Ethyleneimine which the mixed effects model failed to do. However, SLR makes the assumption that all observations within a single study are independent of each other and thus uncorrelated.

While we have conducted a thorough and complete analysis of the data provided, there are some limitations for this current study. Lack of data was a critical aspect within our analysis. All the samples provided had at most six unique observations. While three observations were enough to calculate the ten Berge exponents, this is a rough calculation. It is very dependent on the data and could drastically change with the addition or removal of a single observation. Additionally, this limited data did not allow for us to measure the accuracy of the calculations and our estimates are highly variable. To address this issue, when conducting future studies, it is

advised to have at least three replicates within each study. Additionally, while we were able to compare the estimates between the Rat and Mouse Species for Pentaborane, it would be helpful to expand this out for many other Species and for all of the chemicals to have multiple Species to compare across.

References

Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, www.cdc.gov/.

Dotson, G. Scott, and Richard Niemeier. "Derivation of immediately dangerous to life or health (IDLH) values." (2012).

“Immediately Dangerous to Life or Health.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 10 May 2019, www.cdc.gov/niosh/idlh/default.html.

National Research Council. "Standing Operating Procedures for Developing Acute Exposure Levels for Hazardous Chemicals." (2001).

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

“The Comprehensive R Archive Network - Package ' lme4'.” The Comprehensive R Archive Network, 23 October 2020, cran.r-project.org/web/packages/lme4/lme4.pdf

“Table of IDLH Values.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 8 Oct. 2019, www.cdc.gov/niosh/idlh/intridl4.html.

Ten Berge, W. F., A. Zwart, and L. M. Appelman. "Concentration—time mortality response relationship of irritant and systemically acting vapours and gases." *Journal of Hazardous Materials* 13.3 (1986): 301-309.