Characterization of U-Shaped Exposure-Response Relationships with Application to a Copper Database

by

Brittany Milton

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science in

Statistics

Carleton University
Ottawa, Ontario

©2014
Brittany Milton
Abstract

The need to find an acceptable medium between excess and deficiency motivates the characterization of U-shaped exposure-response relationships. Previous research completed by Chambers et al. (2010) on a copper(Cu) database employed CatReg, a software program developed by the United States Environmental Protection Agency, to model copper excess and deficiency conditions separately. They proposed a single optimal intake level of 2.6 mg Cu/day and an acceptable range of oral intake between 1.8 and 3.1 mg Cu/day. At present, interest lies in the simultaneous fitting of excess and deficiency conditions. Using binary logistic regression as a foundation, we have developed a methodology that can be used to simultaneously fit observations from excess and deficiency experiments. The ability to capture excess and deficiency conditions in one well-defined model facilitates the estimation of a U-shape function, where exposure concentration lies on the horizontal axis and the probability of an adverse health effect lies on the vertical axis. The bottom of this U-shape function represents the dose level of Cu available for intake associated with the smallest chance of adverse health effects attributed to an excess or deficient Cu condition, or both. For humans, this dose level is estimated to be 2.73 mg Cu/day with an acceptable range of oral intake between 1.57 and 4.46 mg Cu/day, which compares very favourably to the results of Chambers et al. (2010).
Acknowledgements

I am very fortunate that I have been able to work with Dr. Patrick Farrell and Dr. Daniel Krewski for these past two years. Working with the two of them has fostered my intellectual development and expanded my horizons. On an individual basis, I would firstly like to thank Dr. Farrell for his continual support and innovative ideas. He is the mastermind behind this thesis and without his guidance, this thesis would have not materialized. I have greatly benefited from his unwavering support and am truly appreciative. Secondly, I would like to thank Dr. Krewski for his patience and encouragement. Over the course of two years, his advice and comments continually improved the quality of this thesis. I would also like to express my gratitude for the financial support.

I would like to thank my family for their endless support, both financially and emotionally, in all of my pursuits. My parents, Ray and Louise, taught me hard work and perseverance by example; the opportunity to learn from them has been the most special invaluable gift I have received. I would also like to thank my boyfriend, Dan, for his kindness and tolerance over these past two years. Without being asked, he always finds some way to help me.
## Contents

1 Introduction ................................................................. 7

2 Literature Review .......................................................... 11
   2.1 Traditional Approaches to Exposure-Response Modeling .... 11
   2.2 Previous Work on the Development of a Model for Copper
       Exposure-Response Data ............................................. 15
       2.2.1 The Proportional Odds Model Applied to Excess and
              Deficiency Data ................................................. 17
   2.3 A New Approach ....................................................... 18

3 Categorical Regression .................................................... 19
   3.1 Categorical Regression Applied to Toxicology Data .......... 19
   3.2 Binary Logistic Regression .......................................... 22
       3.2.1 Parameter Estimation ........................................ 25
   3.3 Ordinal Logistic Regression ......................................... 28
       3.3.1 The Proportional Odds Model .............................. 28
       3.3.2 Parameter Estimation ........................................ 30
   3.4 The Newton-Raphson Method ........................................ 31

4 A Joint Model for Excess and Deficiency .............................. 35
   4.1 Motivation ............................................................. 35
   4.2 Investigating the Point of Intersection ......................... 37
   4.3 Properties of the Estimator for EPCP ........................... 40
   4.4 The Odds of Severity at the EPCP ............................... 43
   4.5 Comparing the Odds of Excess to Deficiency ................... 46
4.6 Stratification by Species ........................................... 47
4.6.1 Humans .......................................................... 49
4.6.2 Mice ............................................................... 49
4.6.3 Rats ................................................................. 50
4.7 Properties of the Estimator for EPCP_{RT}, EPCP_{MU}, EPCP_{HU} 50
4.8 The Odds of Severity at EPCP_{HU}, EPCP_{MU}, and EPCP_{RT} .... 53
4.8.1 Humans .......................................................... 53
4.8.2 Mice and Rats ................................................... 55
4.9 Species-to-Species Odds Ratios ...................................... 59
4.9.1 Mice to Humans and Rats to Humans .......................... 59
4.9.2 Rats to Mice ....................................................... 63
4.10 An Alternative Approach to the Problem : The Independence Model .................................................. 67
4.11 Investigating $x_{\text{MIN, DUE}}$ ........................................ 68
4.12 The Odds of Excess or Deficiency at $x_{\text{MIN, DUE}}$ ............ 71
4.13 The Independence Model Applied to the Species - Stratified JMED ................................................................. 72
4.13.1 Humans .......................................................... 72
4.13.2 Mice ............................................................... 73
4.13.3 Rats ................................................................. 75
4.14 The Odds at $x_{\text{MIN, DUE}}$ for each Species ...................... 76
4.14.1 Humans .......................................................... 77
4.14.2 Mice ............................................................... 77
4.14.3 Rats ................................................................. 78
5 Simulation Study

5.1 Validation Procedure for EPCP ........................................ 79
5.2 Investigating the Limitations of the EPCP .......................... 81
5.3 Validation Procedure for $x_{\text{MIN} \cup \text{E}}$ .......................... 82
5.4 Summary ................................................................. 85

6 Application with Copper Database ........................................ 86

6.1 Representations of the JMED ........................................... 88
6.2 Interpreting the EPCP .................................................. 89
6.3 Interpreting the Odds of Severity at the EPCP ...................... 90
6.4 Odds Ratio of Excess to Deficiency at a Specified $\log_{10}$ (Concentration) ......................................................... 90
6.5 Stratification by Species .................................................. 92
   6.5.1 Interpreting EPCP$_{\text{HU}}$, EPCP$_{\text{MU}}$, and EPCP$_{\text{RT}}$ .... 93
   6.5.2 Interpreting the Log Odds of Severity at EPCP$_{\text{HU}}$, EPCP$_{\text{MU}}$, and EPCP$_{\text{RT}}$ ............................... 95
   6.5.3 Interpreting Species-to-Species Odds Ratios at the EPCP 97
6.6 An Alternative Approach: $x_{\text{MIN} \cup \text{E}}$ ......................... 98
   6.6.1 Interpreting the Odds of Excess or Deficiency at $x_{\text{MIN} \cup \text{E}}$ 99
6.7 The Independence Model Applied to the Species-Stratified
   JMED ........................................................................ 100
   6.7.1 Humans ................................................................ 100
   6.7.2 Mice .................................................................. 102
   6.7.3 Rats .................................................................. 103
7 Application with Manganese Database

7.1 Representations of the JMED .......................... 105

7.2 Interpreting the EPCP ................................. 106

7.3 Interpreting the Odds of Excess or Deficiency at the EPCP . 107

7.4 Stratification by Species ................................. 107
   7.4.1 Interpreting $EPCP_{OTH}, EPCP_{MU}, EPCP_{RT}$ .... 108
   7.4.2 Interpreting the Log Odds at the EPCP ............ 110
   7.4.3 Species-to-Species Odds Ratios at the EPCP ......... 111

7.5 An Alternative Approach: $x_{MIND\cup E}$ ................. 112
   7.5.1 Interpreting the Odds of Excess or Deficiency at $x_{MIND\cup E}$ 113

7.6 The Independence Model Applied to the Species-Stratified
   JMED ..................................................... 114
   7.6.1 Others ............................................. 114
   7.6.2 Mice .............................................. 115
   7.6.3 Rats .............................................. 117

8 Conclusion ................................................. 119
1 Introduction

Copper (Cu) is found in the natural environment in water, soils and fugitive dusts. It is essential in mammalian nutrition as an electron donor or acceptor in metalloenzymes (Ransom Stern et al. 2007). Human exposure to copper primarily stems from the consumption of food and drinking water. Dietary sources of copper are vegetables such as potatoes, mushrooms, and legumes, and certain grains, such as wheat and rye. In humans, the liver is the primary organ of copper-induced toxicity (IPCS 1998). In animals, adverse effects attributed to copper toxicity are observed mainly in organs involved in the absorption and excretion of copper (liver, kidneys, rodent forestomach) (Hebert et al. 1993, IPCS 1998). Other target organs include bone and the central nervous and immune systems (IPCS 1998). In both humans and animals, the major target organs for copper deficiency are the blood and hematopoietic system, the cardiovascular system, connective tissue and bone, the nervous system, and the immune system (Danks 1988, IPCS 1998, Ralph 2001). Recently, the identification of genetic disorders linked to copper metabolism, namely, Wilson disease for copper-induced toxicity, and Menkes disease for copper-induced deficiency, have generated awareness on the potential consequences of copper toxicity in normal and potentially susceptible populations (Ransom Stern et al. 2007). At normal levels, Cu is a natural, harmless constituent in the human body. Because Cu is an essential nutrient, it is possible that both low and high Cu levels could be associated with toxicity (Claus Henn et al. 2010).

The need to find an acceptable medium between copper excess and defi-
ciency motivates exposure-response modeling. Defining the relationship between exposure and the effects the exposure exert on health is necessary in developing health protection programmes for exposure to chemical substances (Ransom Stern et al. 2007). Exposure-response relationships provide the foundation for identifying recommended levels of exposure to essential and nonessential substances (Krewski et al. 2010). The use of a comprehensive database on copper-induced toxicity is essential for exposure-response modeling because a single study does not contain enough information to adequately characterize copper exposure-response relationships (Ransom Stern et al. 2007). Conducting a meta-analysis offers a quantitative method to investigate sources of variation and to reconcile inconsistent results of published observational studies (Il’yasova et al. 2005). In 2002, a group of toxicologists, biologists, and epidemiologists formed a working group to review the literature on Cu excess and deficiency; the information gathered from this collaborative effort was organized into a database comprised of multiple studies examining a wide range of adverse health effects attributed to excess and deficiency. Integrating multiple studies that measure outcomes in different target organs with varying levels of severity fostered the development of a common response scale. To create a common measure of response, potential health effects emerging from excess or deficient conditions were organized into ordered categories of severity, then each experiment was assigned a unique severity score. This procedure standardized the studies with respect to severity of adverse health outcomes (Ransom Stern et. al 2007). This database has been updated to include studies published between 2002 - 2008 and includes information extracted from each study, namely, animal
species, exposure medium, Cu species, age, gender, study design, and dose and duration. In total, there were 42 excess studies and 79 deficiency studies.

Prior to current research, the absence of a well-defined exposure-response curve impeded the recommendation of a safe Cu intake level. Ransom Stern et al. (2007) suggested the dose-response curve likely displayed a U-shape, since both excess and deficiency produce adverse health effects. This U-shaped response curve occurs when excess and deficiency curves are plotted on a continuum (x axis) ranging from very low doses to very high doses. The probability of severity (y axis) increases with very low or very high doses (Chambers et al. 2010). Ransom Stern et al (2007) propose this U-shape curve is asymmetrical. Developing an expression for this U-shape curve requires robust methods that capture the differential behaviour exhibited by the excess and deficiency curves. Current research has verified this claim and endeavors to identify the minimal point on this curve, namely, the bottom of the U. From a health-related perspective, this point is significant because it represents the recommended Cu intake associated with the lowest probability of experiencing adverse health effects attributed to an excess Cu condition, a deficient Cu condition, or both. Furthermore, a well-defined curve provides the opportunity to stratify by species, perform inference, derive confidence limits and validate results using computing techniques.

In this thesis, we develop a joint model based on logistic regression for simultaneously describing copper excess and deficiency, which allows for the development of separate excess and deficiency curves. In Chapter 2, we discuss the traditional approaches to exposure-response modeling and present a literature review of the research conducted on copper data. Chapter 3
begins with an introduction to binary logistic regression and the procedures for obtaining parameter estimates for the associated model, along with an estimated asymptotic covariance matrix. Given that copper excess and deficiency data are often measured on an ordinal scale, we also describe ordinal logistic regression modeling for accommodating this type of data. To address the problem of simultaneously modeling the excess and deficiency curves, we explore two different approaches which we present in Chapter 4. For simplicity, we initially consider only one continuous covariate and dichotomize copper outcomes so that binary logistic regression may be used. The first approach is based on the intersection point of the excess and deficiency curves arising from the joint modeling approach. The second approach imposes an independence assumption on excess and deficiency outcomes, and considers the point at which the probability of an excess(E) or deficient(D) condition, or both(D∪E) is minimized. Inferential techniques for the assessment of confidence intervals based on these two approaches are developed. Extensions to more than one covariate are also considered; for example, an illustration where responses are stratified by species is presented. In Chapter 5, we present the results of a simulation study designed to assess the performance of the proposed techniques in Chapter 4, and discuss the limitations of these approaches. In Chapter 6, we demonstrate an application of the methodology developed in Chapter 4 to a copper database. Since the proposed approaches are more generally applicable, we illustrate a second example in Chapter 7 using a manganese database. Finally, conclusions and discussions are provided in Chapter 8.
2 Literature Review

2.1 Traditional Approaches to Exposure-Response Modeling

The notion of identifying a ‘safe’ dose in non-cancer risk assessment has been investigated by expert risk assessors worldwide (Lu et al., 1991). Health risk assessments performed by regulatory agencies have one of two goals: to estimate an exposure level where health risk is minimal or zero, or to estimate the health risk at existing or proposed exposure levels (Hertzberg and Dourson, 1993). The concept of a Reference Dose, developed by the United States Environmental Protection Agency (US EPA), has been accepted as the main strategy for addressing the first objective. No procedures for non-cancer risk have been universally adopted by the US EPA for the second objective. When human data are available, there are three traditional approaches that have been used for estimating human risk.

The most widely accepted approach involves an expert panel judgment of risk proposing a Reference Dose (RfD). The US EPA defines the RfD as “an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily oral exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime.” The RfD relies heavily on expert judgment, in fact, a judgment-based decision is implemented in several steps of the determination process (Hertzberg and Dourson, 1993). First, a literature review must be performed to determine the critical effect under study. Second, the most
scientifically sound study with the most appropriate NOAEL (No Adverse Effects Level) is selected. The NOAEL is concerned with identifying ‘safe’ or ‘acceptable’ intake levels (Edler et al, 2002). Finally, the uncertainty factors must be assigned. In practice, uncertainty factors have been used to extrapolate the NOAEL from an animal experiment dose level to a human safe dose level based on the unknown (Teuschler et al., 1999). Simply stated, these uncertainty factors account for interspecies and/or intraspecies differences. Traditionally, toxicologists use an uncertainty factor of 10. The RfD is obtained by dividing the NOAEL by one or more of the uncertainty factors (Hertzberg and Dourson, 1993).

The RfD has been used as the traditional approach for human risk estimation because it is based on unanimous expert group judgment and it considers all available data corresponding to the selected critical toxic effect in the estimation process (Hertzberg and Dourson, 1993). Despite the expert judgment built into the RfD, the procedure for calculating the RfD does not include any measure of the estimation of error or the variability in the RfD. This is problematic because expressing the potential consequences of excess exposures is not possible (Teuschler et al., 1999). For excess exposures, it is possible that dividing a NOAEL by an uncertainty factor may shift the daily intake value into the deficiency range where health is no longer maintained. Furthermore, there would be a range of copper values when deficiency and excess might occur in different individuals at the same copper level (Ransom Stern et al., 2007).

By definition, the RfD implies that an exposure at that level corresponds
to a minimal or zero risk of adverse health effects. The RfD is not interpreted as a threshold, rather, RfD’s are believed to be near the upper end of the range of acceptable exposure levels (Hertzberg and Dourson, 1993). This interpretation can be transformed into a probabilistic, mathematical definition. To mitigate some of the issues associated with the RfD, toxicologists introduced Bayesian modeling of the critical effect to obtain a benchmark dose (BMD). The BMD is a modeled value defined by the U.S. EPA as “the statistical lower confidence limit on the dose estimated to produce a predetermined level of change in response relative to controls” (Teuschler et al., 1999). A preset benchmark risk is selected by fitting a mathematical model to quantal dose-response data based on one critical effect at one time point in a single study. This benchmark risk level is chosen to be low and is often interpreted as the lowest incidence that is statistically significant in common toxicology studies (Hertzberg and Dourson, 1993). The estimate of the BMD is obtained by finding the corresponding dose that satisfies the preselected benchmark risk level. The RfD is calculated by replacing the NOAEL with the BMD, so that the RfD is simply the quotient of the BMD and the uncertainty factors.

Using the BMD instead of the NOAEL improves the RfD estimation because it considers data from the entire dose-response curve for the selected critical effect, rather than only from the single dose (Edler et al, 2002). The BMD approach builds a model based on the information contained within the range of the study. Extrapolation to low doses is not recommended. Modeling of the BMD is a curve fitting exercise lacking any biological basis (Haber et al., 2001). Crump (1984) demonstrated the uncertainty associated
with the BMD; provided model fit is adequate, any model can be used to find an estimate for the BMD. Extrapolating beyond the data increases model dependence. If a variety of models can adequately describe the range of the data, these models will likely have widely different results at doses several orders of magnitude lower. In effect, it is possible to obtain numerous BMD estimates that all emerge from a ‘good’ model fit.

Integrating a mathematical model into the estimation procedure affords the opportunity to justify the selection of the estimate. This is the benefit of using the BMD rather than the NOAEL. The performance of the BMD as a risk estimator is hindered by its’ inability to use all of the data available (Hertzberg and Dourson, 1993). A single critical toxic effect from one particular study is selected for the modeling procedure. This does not include sufficient amounts of information to make any solid generalizations regarding risk estimates for humans. BMD modeling can not account for several adverse health effects that may occur simultaneously (Birkett et al., 2010). Furthermore, the BMD method does not account for the judgment of adversity or the selection of uncertainty factors.

The complex toxicological properties of essential metallic elements is a challenging prospect for traditional risk assessment modeling approaches. Categorical regression may be considered the most appropriate risk assessment tool to define an acceptable oral range of intake since it has the ability to incorporate all available data on the toxic effects associated with such elements (Ransom Stern et al., 2007).
2.2 Previous Work on the Development of a Model for Copper Exposure-Response Data

Previous work with the copper database considered here was completed by Chambers et al. (2010). For information on the assembly of this database, please refer to ‘Development of a Copper Database for Exposure-Response Analysis’ by Chambers et al. (2010) and Ransom Stern et al.’s (2007) paper. The authors used CatReg, a software program developed by the US EPA, to perform a categorical regression analysis using the copper database and created separate models for excess and deficiency. Chambers et al. (2010) found the exposure response curves for copper excess and deficiency defined an optimal intake level of 2.0 mg Cu/day and an acceptable range of oral intake (AROI) between 1.8 and 3.1 mg Cu/day. In 2001, the Food and Nutrition board recommended 0.9 mg Cu/day for adult men and women, and suggested an upper safe limit of 10 mg Cu/day.

The above analysis makes use of group level data. Specifically, a single record in the database corresponds to one dose group, rather than one individual. To account for the possible differences in group sizes, a weight factor was assigned to each dose group in the database; see Table 1 for a summary. Most studies used in the database reported multiple responses to copper excess or deficiency, resulting in several severity scores associated with one exposure level. Severity scores ranged from 0 to 6, with 0 being least severe and 6 being most severe. Severity scores were assigned by isolating each independent endpoint and assigning one severity score to that outcome. The most severe effect was selected and used to represent that exposure group.
In this approach, there is an underlying assumption that the mean response is representative of the sample of subjects in the exposure group, irrespective of whether they experienced the event. In effect, the probability of severity obtained from the regression model represents the likelihood that one would assign the entire dose group to a given severity level or greater (Hertzberg and Dourson, 1993).

<table>
<thead>
<tr>
<th>Dose Group Size</th>
<th>Weight Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 10</td>
<td>1</td>
</tr>
<tr>
<td>11 - 25</td>
<td>2</td>
</tr>
<tr>
<td>≥ 26</td>
<td>3</td>
</tr>
</tbody>
</table>

Groups of observations from the same reference and experiment were treated as clusters. CatReg assumes these clusters are correlated, while observations from different clusters are independent. CatReg uses the method of generalized estimating equations to account for the cluster sampling effect (Simpson et al., 1996; Diggle et al. 2002). The analysis assumes highly correlated responses in individuals within a group and treats groups as individuals with an effective sample size of n=1 (Guth et al, 1997).

In the analysis, severity level 2 was assigned to responses associated with early responses to excess or deficient levels of copper. Chambers et al. (2010) used CatReg to generate estimates for the 10th percentile of the extra risk concentration, ERC\textsubscript{10}, at severity level 2 or greater for the human stratum with the copper excess and deficiency data sets. These were the values used to define the limits of the AROI.

Mathematically, the ERC\textsubscript{q} is a concentration c, at time t, defined as:
\[
P (Y \leq s | C = c, T = t) - P (Y \leq s | C = 0, T = t) \over 1 - P (Y \leq s | C = c, T = t) = \frac{q}{100}
\]

for q between 1 and 100. In other words, the ERC_q at time t is the exposure concentration at which the probability is \( \frac{q}{100} \) of an adverse effect of level s or higher.

CatReg offers two models for ordinal severity data: the proportional odds model, which is also referred to as the cumulative odds model, and the unrestricted cumulative model. The cumulative odds model is a general model, while the unrestricted cumulative model is more complex. First, excess and deficiency data were modeled separately with the proportional odds model.

### 2.2.1 The Proportional Odds Model Applied to Excess and Deficiency Data

The proportional odds model defines an ERC_{10} estimate for severity level 2 or greater for copper deficiency (2.3 mg/day) that is higher than the ERC_{10} estimate for copper excess (1.8 mg/day). When these curves are superimposed onto the same window, they cross at approximately 2.1 mg Cu/day. Using the two-level model, the midpoint between the ERC_{10} for copper deficiency and copper excess would have been approximately 2.0 mg Cu/day. Note that the unrestricted cumulative odds model was also considered, but results were not used in further investigation. The Cu database was updated to include all published studies between 2002 and 2008. Using the updated database, Chambers et al. (2010) found the single optimal intake level to be 2.6 mg/day.
2.3 A New Approach

CatReg was designed to model acute inhalation exposure data, which is characterized by an increasing relationship. In light of this designation, CatReg understands a decreasing relationship as evidence of no effect and generates an error message. To successfully use CatReg for the deficiency analysis, Chambers et al. (2010) had to impose an increasing relationship by adjusting the response variable. This adjustment made it impossible to simultaneously model excess and deficient conditions with CatReg.

Interest lies in finding a single optimal intake level of copper where there is a small risk of a non-normal reading caused by an excess condition, a deficient condition, or both. To accomplish this feat, it is necessary to find a functional form for a joint model whose components consist of an excess model and a deficiency model that are fitted simultaneously. CatReg is unable to perform this function since it is limited to fitting one exposure-response model at a time.

In this study, we consider the use of logistic regression to simultaneously model both excess and deficient conditions. For simplicity, we restrict our investigation to non-clustered binary data. However, we would expect that the techniques developed should extend to more complex data, including those measured on an ordinal scale.
3 Categorical Regression

3.1 Categorical Regression Applied to Toxicology Data

Categorical regression is a mathematical tool that can estimate potential health risks from chemical exposures. The application of categorical regression methods to health risk assessment was proposed by Hertzberg and Miller (1985) and Hertzberg and Dourson (1993) to empirically evaluate differences in sensitivity for toxicity data with ordered responses. A collection of studies is examined and various critical effects are recorded. These critical effects are assigned to ordinal severity categories (e.g., no effect, mild effect, adverse effect). The regression model builds a relationship describing the ordinal response scores in terms of the exposure information (Haber et al., 2001). In performing a regression on ordered categories of toxic severity, this method can estimate the likelihood that a given category of severity may occur. This model does not suggest you directly observe the effect (Edler et al., 2002).

Categorical regression is appealing to risk assessors because it facilitates the use of multiple independent variables to explain a given response. Most commonly, concentration and duration are included in the model for analysis. Using the regression model, it is possible to predict the response based on controlling the effects of all covariates (Haber et al., 2001). This model flexibility provides ample opportunity for analysis.

Categorical regression has been proposed as a “superior alternative” to traditional approaches of dose-response modeling because it addresses the obstacles encountered by the RfD and BMD (Hertzberg and Dourson 1993). There are two obstacles impeding the performance of the RfD and BMD
as risk estimates. The first is their inability to use all available data on
the critical effects exhibited by a chemical. The second is the absence of a
method for expressing the uncertainty in these estimates. The major advan-
tage of using categorical regression is all relevant studies that report differ-
ent critical effects are incorporated. Classifying critical effects into severity
categories creates a system where different endpoints are standardized on
a common severity scale. Ultimately, this enables all available data to be
used in the analysis (Haber et al., 2001). Combining multiple studies in
the meta-analysis dramatically increases the statistical power of the anal-
ysis. Estimates obtained from the RfD or BMD approaches are based on
limited amounts of information whereas results generated from a categorical
regression are based on all available information. The second obstacle asso-
ciated with the RfD and BMD is also resolved; the uncertainty in statistical
regression lies in the possibility that the toxicologist would assign a differ-
ent category to an observed effect. The uncertainty is implemented within
the model and is directly interpreted in terms of judgment (Hertzberg and
Dourson 1993).

In risk assessment, data sets often contain gaps where there is no informa-
tion available for a bracketed time or concentration interval. To address these
inconsistencies, traditional methods assigned uncertainty factors to obtain
risk estimates. When using categorical regression, all information pertaining
to the explanatory capacity of the duration, concentration, and dosage in
predicting severity is retained within the model. This provides the opportu-
nity for interpolation or extrapolation. Using the fitted model, it is possible
to predict the severity of the effect for a bracketed time or concentration
interval for which there are no measurements by using related information at different durations, species, or exposure levels (Edler et al., 2002; Haber et al., 2001).

Although categorical regression is thought to be the most useful tool available for toxicology analysis (Ransom Stern et al., 2007), criticisms still emerge regarding its application. Considerable toxicological judgment is needed to rate the severity of the effect and to uniformly categorize the effects across multiple studies. This introduces subjectivity into the process (Edler et al., 2002). In addition to subjectivity, categorization results in a loss of information about the magnitude and variability of responses, however, it is necessary to accommodate the standardization of continuous, categorical, and descriptive data (Guth et al., 1997). To elaborate on this point, response categories are ordered by severity, with spacing among the ordinal scores having no direct interpretation. For example, a score of 2 is higher than 1, but the difference in effects may not be the same as the difference between 3 and 2 (Haber et al., 2001). The categorical regression model provides no way of capturing differences that may exist among the qualitatively-ordered severity effects. Furthermore, a good model fit indicated by a goodness of fit test should not be extended to include extrapolated data. Although a model may appear to fit the data well, there is no way to evaluate the accuracy achieved in extrapolation (Edler et al., 2002). Finally, categorical regression is statistically driven; there is no information on biological processes involved in developing the model. Implementing a mechanism that integrates biological reasoning into the model would provide a statistically and biologically plausible method (Krewski et al., 2010).
The application of categorical regression requires a special type of response variable, \( Y \). In the general case, this response variable is comprised of \( S \) ordered categories. This response variable may be continuous, descriptive, or categorical, provided the effects are classified into ordered categories (Haber et al., 2001). In developing the concept of a categorical regression, it is helpful to begin with the special case whereby the response falls into one of two categories with or without order being assumed.

### 3.2 Binary Logistic Regression

Consider a random variable \( Y \) that can take on one of two possible values, 0 or 1. Given a dataset of \( n \) such independent observations, then \( Y \) can be considered a column vector of \( n \) Bernoulli random variables. Agresti (1990) developed the foundation of binary logistic regression models. By convention, for the \( i^{th} \) observation \((i=1, \ldots, n)\) Agresti has denoted \( Y_i = 1 \) representing a success and \( Y_i = 0 \) representing a failure.

In performing any regression, there is an underlying assumption that the response variable is related to one or more of \( K \) explanatory variables, \( x_{i1}, \ldots, x_{iK} \), according to a specified functional relationship. For the case of the Bernoulli random variable, \( Y_i \), above, we model its associated parameter, \( \pi_i \), as \( P(Y_i = 1|x_{i1}, \ldots, x_{iK}) = \pi_i \) and \( P(Y_i = 0|x_{i1}, \ldots, x_{iK}) = 1 - \pi_i \), and introduce a link function for \( \pi_i \). The link function is often taken to be the logistic function since it is easily converted to a linear relationship by the logit transformation (Dourson et al., 1997). For a binary response variable, another option is the probit link function, in which \( \pi_i \) is modeled according to
the cumulative distribution of the standard normal distribution. Neverthe-
less, the focus here will remain on the logistic function since it is commonly
used in practice.

The logistic function defines $P(Y_i = 1|x_{i1}, \ldots, x_{iK})$ as

$$
\pi_i = \frac{\exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})}{1 + \exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})}
$$

(3.1)

where $\pi_i \in (0, 1)$ and $\beta^T = [\beta_0, \beta_1, \ldots, \beta_K]$ is a vector of regression param-
eters consisting of an intercept term, $\beta_0$, and slope parameters $\beta_0, \beta_1, \ldots, \beta_K$.
It is necessary to constrain the components in the vectors $x_1, \ldots, x_K$ to be
non-unique.

Consequently, $P(Y_i = 0|x_{i1}, \ldots, x_{iK})$ is given by

$$
1 - \pi_i = 1 - \frac{\exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})}{1 + \exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})}
$$

$$
= \frac{1 + \exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})}{1 + \exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})} - \frac{\exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})}{1 + \exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})}
$$

$$
= \frac{1}{1 + \exp(\beta_0 + \beta_1x_{i1} + \cdots + \beta_Kx_{iK})}
$$

(3.2)
To linearize $\pi_i$, the logit transformation may be applied:

$$
\text{logit}(\pi_i) = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) \\
= \ln \left( \frac{\exp (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK})}{1 + \exp (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK})} \cdot \frac{1 + \exp (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK})}{1} \right) \\
= \ln \left( \exp (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK}) \right) \\
= \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK}
$$

when $x_{i0} = 1$ for $i = 1, \ldots, n$.

The linear component of the model contains the design matrix and the vector of parameters to be estimated. The design matrix, $X$, is comprised of $n$ rows and $K+1$ columns. The first column in the design matrix is a vector of 1’s corresponding to the intercept term, $\beta_0$. The remaining columns are the $K$ independent variables, $(x_{i1}, x_{i2}, \ldots, x_{iK})$.

$$
X_{n\times(K+1)} = \begin{bmatrix}
1 & x_{i1} & \cdots & x_{iK} \\
1 & x_{21} & \cdots & x_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & \cdots & x_{nK}
\end{bmatrix}
$$

(3.4)

The logistic regression model equates the logit transform to the linear component:

$$
\text{logit}(\pi_i) = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) = \sum_{k=0}^{K} \beta_k x_{ik} = \beta^T X
$$

(3.5)
3.2.1 Parameter Estimation

The goal of binary logistic regression is to estimate the $K+1$ unknown parameters in (3.3). To achieve this, maximum likelihood estimation is employed. Maximizing the likelihood entails finding the set of parameters for which the probability of the observed data is greatest (Agresti 1990). The maximum likelihood equation is derived from the probability distribution of the dependent variable, $Y$. Since each $y_i$ represents a Bernoulli random variable for the $i^{th}$ observation, the distribution of the data, or likelihood, is:

$$L(\beta | y) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$ (3.6)

where $\pi_i$ is defined in (3.1).

The maximum likelihood estimates (MLE) are the values for $\beta$ that maximize the likelihood function in (3.6). The critical points of a function (maxima and minima) occur when the first derivative equals zero. If the second derivative evaluated at that point is less than zero, then the critical point is a maximum. Finding the MLE requires computation of the first and second derivatives of the likelihood function. This procedure is simplified by using the log-likelihood function, which is simply attained by applying the natural log transform to (3.6), giving

$$l = \log L = \sum_{i=1}^{n} y_i \log (\pi_i) + \sum_{i=1}^{n} (1 - y_i) \log (1 - \pi_i)$$ (3.7)

Maximality will be preserved since the logarithm is a monotonic function. To develop the MLE, we must differentiate (3.7) with respect to $\beta_j$ for $j =$
\[ \frac{\partial l}{\partial \beta_k} = \sum_{i=1}^{n} y_i \frac{\partial \log (\pi_i)}{\partial \beta_k} + \sum_{i=1}^{n} (1 - y_i) \frac{\partial \log (1 - \pi_i)}{\partial \beta_k} \] (3.8)

Note:

\[ \frac{\partial \log (\pi_i)}{\partial \beta_k} = \frac{\partial \log (\pi_i)}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \beta_k} = \frac{1}{\pi_i} \cdot \frac{\partial \pi_i}{\partial \beta_k} \] (3.9)

\[ \frac{\partial \log (1 - \pi_i)}{\partial \beta_k} = \frac{\partial \log (1 - \pi_i)}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \beta_k} = - \frac{1}{1 - \pi_i} \cdot \frac{\partial \pi_i}{\partial \beta_k} \] (3.10)

Thus,

\[ \frac{\partial \pi_i}{\beta_k} = \frac{(1 + \exp(\sum_{k=0}^{K} \beta_k x_{ik})) x_{ik} \exp(\sum_{k=0}^{K} \beta_k x_{ik}) - \exp(\sum_{k=0}^{K} \beta_k x_{ik}) x_{ik} \exp(\sum_{k=0}^{K} \beta_k x_{ik})}{(1 + \exp(\sum_{k=0}^{K} \beta_k x_{ik}))^2} \] (3.11)

Let

\[ Z = \sum_{k=0}^{K} \beta_k x_{ik} \]

Now we may rewrite (3.11) as

\[ \frac{\partial \pi_i}{\beta_k} = \frac{x_{ik}[(1 + \exp(Z)) \exp(Z) - \exp(Z) \exp(Z)]}{(1 + \exp(Z))(1 + \exp(Z))} \]

\[ = \frac{x_{ik} \exp(Z)}{(1 + \exp(Z))(1 + \exp(Z))} \]

\[ = x_{ik} \pi_i (1 - \pi_i) \] (3.12)

Substituting (3.9), (3.10), and (3.12) into (3.8), we obtain
\[
\frac{\partial l}{\partial \beta_k} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} x_{ik} \pi_i (1 - \pi_i) - \sum_{i=1}^{n} \frac{(1 - y_i)}{(1 - \pi_i)} x_{ik} \pi_i (1 - \pi_i) \\
= \sum_{i=1}^{n} x_{ik} y_i - \sum_{i=1}^{n} x_{ik} \pi_i
\] 

(3.13)

To find the MLE, set \( \frac{\partial l}{\partial \beta_k} = \sum_{i=1}^{n} x_{ik} y_i - \sum_{i=1}^{n} x_{ik} \pi_i = 0 \). These MLEs are nonlinear in \( \beta \); however, they can be solved using a Newton Raphson Algorithm, as described in Section 3.4. The variance-covariance matrix associated with the estimators of \( \beta \) is simply the negated inverse of the information matrix, denoted as \( \mathcal{I}(\beta) \), which takes the general form:

\[
\mathcal{I}(\beta) = 
\begin{bmatrix}
\frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_1} & \frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_K} \\
\frac{\partial^2 l(\beta)}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 l(\beta)}{\partial \beta_2 \partial \beta_2} & \cdots & \frac{\partial^2 l(\beta)}{\partial \beta_2 \partial \beta_K} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 l(\beta)}{\partial \beta_K \partial \beta_1} & \frac{\partial^2 l(\beta)}{\partial \beta_K \partial \beta_2} & \cdots & \frac{\partial^2 l(\beta)}{\partial \beta_K \partial \beta_K}
\end{bmatrix} 
\] 

(3.14)

where

\[
\frac{\partial^2 l(\beta)}{\partial \beta_m \partial \beta_k} = -\frac{\partial}{\partial \beta_k} \left( \sum_{i=1}^{n} x_{im} \pi_i \right) \\
= -\sum_{i=1}^{n} x_{im} \frac{\partial \pi_i}{\partial \beta_k} \\
= -\sum_{i=1}^{n} x_{im} x_{ik} \pi_i (1 - \pi_i)
\] 

(3.15)
3.3 Ordinal Logistic Regression

McCullagh (1980) considered the general case when the response variable is comprised of \( S \) ordered categories. These categories are assigned to a scheme that qualitatively define different levels of an effect (e.g. no effect, low effect, high effect, etc.). The categories are ordered such that an increase in category reflects an increase in severity of the effect. We will examine a specific case where the \( S - 1 \) cumulative logit functions for the \( S \) severity categories differ only by their intercept value.

3.3.1 The Proportional Odds Model

McCullagh (1980) developed the Proportional Odds Model. The overarching theme of the Proportional Odds Model is that the difference between corresponding cumulative logits is independent of the severity category involved. In effect, the \( S - 1 \) cumulative logits corresponding to the \( S \) severity categories differ only by their intercept value.

Let \( \kappa_j(x) \) be the odds that the severity category for the \( i^{th} \) observation is less than or equal to some severity category \( j \), that is \( Y_i \leq j \), given the covariate values \( x \). Then the proportional odds model specifies that:

\[
\kappa_j(x) = \kappa_j \exp(\beta'x)
\]

By definition, the odds may be expressed as:

\[
\kappa_j(x) = \frac{\gamma_j}{1 - \gamma_j} \text{ where } \gamma_j = \sum_{k=1}^{j} \pi_k = \pi_1 + \pi_2 + \cdots + \pi_j
\]
Then the expression for the odds may be rewritten as:

$$\frac{\gamma_j}{1 - \gamma_j} = \kappa_j \exp(\beta'x)$$

By applying the logarithmic transform, the logit transform is indirectly applied:

$$\log \left( \frac{\gamma_j}{1 - \gamma_j} \right) = \log [\kappa_j \exp(\beta'x)] = \log \kappa_j + \beta'x$$

so the proportional odds model is identical to the linear logistic model:

$$\log \left[ \frac{\gamma_j}{1 - \gamma_j} \right] = \alpha_j + \beta'x \quad (1 \leq j < s) \quad \text{with} \quad \alpha_j = \log \kappa_j.$$

Now, the S - 1 cumulative logits may be represented as:

$$L_1 = \log \left( \frac{\pi_1}{1 - \pi_1} \right) = \log \left( \frac{\pi_1}{\pi_2 + \cdots + \pi_s} \right) = \alpha_1 + \beta'x$$

$$L_2 = \log \left( \frac{\pi_2}{1 - \pi_2} \right) = \log \left( \frac{\pi_1 + \pi_2}{1 - \pi_2} \right) = \log \left( \frac{\pi_1 + \pi_2}{\pi_3 + \pi_4 + \cdots + \pi_s} \right) = \alpha_2 + \beta'x$$

$$\vdots$$

$$L_{S-1} = \log \left( \frac{\pi_{S-1}}{1 - \pi_{S-1}} \right) = \log \left( \frac{\pi_1 + \cdots + \pi_{S-2}}{\pi_s} \right) = \alpha_{S-1} + \beta'x$$

For example, the log of the odds that a given observation will be less than or equal to severity category h, is $\alpha_h + \beta'x$. $\beta_{[K+1] \times [S-1]}$ is a matrix where each element, $\beta_{k,j}$, contains the parameter estimate for the $k^{th}$ covariate and the $j^{th}$ value of the dependent variable, where $k = 1, \ldots K$ and $j = 1, \ldots S - 1$. The entries in the first row of $\beta$ are the intercept terms that
vary by severity category, however, the remaining rows remain constant for all severity categories. The design matrix, $X$, remains the same as (3.4).

### 3.3.2 Parameter Estimation

Consider a random variable $Y$ that can take on one of $S$ possible severity outcomes. Let $n_j$ represent the number of observations in the $j^{th}$ class of the dependent variable. Then $Y \sim \text{multinomial}(n_1, \ldots, n_s)$ where $n_j$ represents the proportion of the observations in severity category $j$ and $\sum_{j=1}^{s} n_j = 1$. $Y$ is a matrix with $n$ rows and $S-1$ columns where each entry in $Y$, $y_{ij}$, represents the observed counts of the $i^{th}$ observation in the $j^{th}$ severity category. Consequently, $\pi$ is a matrix with dimension $n \times (S - 1)$, where the entry $\pi_{ij}$ represents the probability of observing the $j^{th}$ value of the dependent variable for the $i^{th}$ observation.

Define

$$
\begin{align*}
R_1 &= n_1 \\
R_2 &= n_1 + n_2 \\
R_3 &= n_1 + n_2 + n_3 \\
&\vdots \\
R_S &= 1
\end{align*}
$$

The joint probability of $(Y_1, Y_2, \ldots, Y_n)$ is proportional to the product of $n$ multinomial functions. Since we are dealing with cumulative probabilities,
The likelihood can be written as a product of the $S - 1$ quantities.

The likelihood function may be written as

$$\mathcal{L}(\beta|Y) = \prod_{i=1}^{n} \left[ \left( \frac{\gamma_1^{(i)}}{\gamma_2^{(i)}} \right) \cdot \left( \frac{\gamma_2^{(i)} - \gamma_1^{(i)}}{\gamma_2^{(i)}} \right) \cdot \left( \gamma_2^{(i)} - \gamma_1^{(i)} \right) \cdot \left( \frac{\gamma_2^{(i)} - \gamma_1^{(i)}}{\gamma_2^{(i)}} \right) \right]$$

(3.16)

The MLEs may be obtained by computing the first and second derivatives of the likelihood function. Applying the natural log transform to (3.16), we obtain the log-likelihood function. We must differentiate the log-likelihood function with respect to $\beta_{k,j}$ for $k = 1, \ldots K$ and $j = 1, \ldots S - 1$. Using the same approach employed for binary logistic regression, we can develop the MLE and its associated information matrix.

### 3.4 The Newton-Raphson Method

Differentiating the log-likelihood equations associated with the binomial and multinomial logistic regression models with respect to $\beta$ results in a system of $K+1$ nonlinear equations with $K+1$ unknown variables and $(S-1) \cdot (K+1)$ nonlinear equations with $(S-1) \cdot (K+1)$ unknown variables, respectively.

Solving a system of nonlinear equations requires numerical estimation using an iterative process. A common method used is the Newton-Raphson Algorithm, also called Newton’s Method. For simplicity, we will consider the case where $\beta$ is a scalar quantity; the results readily extend to the vector or matrix case. This method begins with an initial guess, $\beta^{(0)}$, and uses the first two terms of the Taylor polynomial evaluated at $\beta^{(0)}$ to come up with another estimate that is closer to the solution.
Assuming the first \( n \) derivatives of \( f = l(\beta) \) exist at \( \beta^{(0)} \), the Taylor polynomial of degree \( n \) for \( f \) at the point \( \beta = \beta^{(0)} \) is defined as the first \( n \) terms of the Taylor series for \( f \):

\[
f^{(n)}(\beta^{(0)}) = \sum_{i=0}^{n} \frac{f^{(i)}(\beta^{(0)})}{i!} (\beta - \beta^{(0)})^i
\]  

(3.17)

The first degree Taylor polynomial is:

\[
f^{(1)}(\beta^{(0)}) = f(\beta^{(0)}) + f'(\beta^{(0)}) \cdot (\beta - \beta^{(0)})
\]  

(3.18)

(3.18) is the equation for the tangent line to \( f \) at \( (\beta^{(0)}, f(\beta^{(0)})) \). Setting this expression equal to zero and solving for \( \beta \) gives:

\[
\beta = \beta^{(0)} - \frac{f(\beta^{(0)})}{f'(\beta^{(0)})}
\]  

(3.19)

This new value for \( \beta \) is the next approximation for the root. Setting \( \beta^{(1)} = \beta \) and repeating this algorithm, this process continues until it converges to the solution. To verify \( \beta \) is the solution, one must check that the matrix of second partial derivatives is negative definite, which guarantees the solution is a global maximum rather than a local maximum.

Newton’s method can be generalized to the estimation of more than one parameter, where we consider a parameter vector \( \beta' = (\beta_0, \ldots, \beta_k) \).

In matrix notation, the first step, (3.18), may be expressed as:

\[
\beta^{(1)} = \beta^{(0)} + \left[ -f''(\beta^{(0)}) \right]^{-1} \cdot f'(\beta^{(0)})
\]  

(3.20)
Define $\mu$ as a column vector of length $N$ with elements $\mu_i = n_i \pi_i$. Using matrix multiplication, it is possible to show that

$$f'(\beta) = X^T(y - \mu)$$ \hspace{1cm} (3.21)

For the binary logistic regression model, $\mu$ is a column vector of length $K + 1$. For the multinomial logistic regression model, $\mu$ is a matrix with dimension $N \times (S - 1)$.

Let $W$ be a square matrix of order $N$ with elements $n_i \pi_i (1 - \pi_i)$ on the diagonal and zeros elsewhere. Using matrix multiplication, it is possible to verify that

$$f''(\beta) = -X^T WX$$ \hspace{1cm} (3.22)

(3.22) generates a $(K+1) \times (K+1)$ square matrix. For the multinomial case, when $j = j'$, $W$ is a a matrix of order $N$ with diagonal elements $n_i \pi_{ij} (1 - \pi_{ij})$ and zeros elsewhere. This formulation can only be used for the diagonal elements. For the off-diagonal elements where $j \neq j'$, $W$ is a diagonal matrix with elements $n_{i,j'j}p_{ik}$; (3.22) should be negated.

Now, (3.18) may be rewritten as

$$\beta^{(1)} = \beta^{(0)} + [X^T WX]^{-1} \cdot X^T(y - \mu)$$ \hspace{1cm} (3.23)

For convergence, we choose to iterate with (3.23) until $||\beta^{(k)} - \beta^{(k-1)}|| < \epsilon$, for a sufficiently small $\epsilon$. At this point, the maximum likelihood estimates
of $\beta$ have converged. The associated variance-covariance matrix is given by (3.22).
4 A Joint Model for Excess and Deficiency

4.1 Motivation

In this chapter, we propose a logistic regression model that permits the simultaneous modeling of both the excess and deficiency responses. For simplicity, it is assumed that there is only one continuous covariate, namely, the logarithm(base 10) of concentration. In addition, we also assume that the responses have been dichotomized so that logistic regression may be used.

The Joint Model for Excess and Deficiency (JMED) is defined to capture the two most informative factors, dose concentration and origin of toxicity (excess or deficient), in one well-defined model. The use of an indicator variable to inform whether an observation originates from the excess or deficiency data set ultimately allows for the creation of an asymmetrical U-shape curve, which reflects the differential behaviour observed in the deficiency and excess data sets. The JMED model provides a representation of exposure-response data where all information is contained within one model; however, it also affords the opportunity to express excess and deficiency curves separately. The primary focus is to find the intersection point of the excess and deficiency curves.

Define

\[ x_{i1} = \log_{10} \text{concentration of the } i^{th} \text{ observation} \]
The JMED model represents the probability of a severe reading as:

\[
P(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2})}
\]  

(4.1)

The probability of severity from an excess reading is:

\[
P_E(Y_i = 1) = \frac{\exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}]}{1 + \exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}]}
\]  

(4.2)

The probability of severity from a deficiency reading is:

\[
P_D(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 x_{i1})}{1 + \exp(\beta_0 + \beta_1 x_{i1})}
\]  

(4.3)

Specifically, the copper database to be examined is such that each trial is linked to a study either focusing on excess or deficiency. However, an observation linked to a given study does not necessarily exhibit the condition of interest in that investigation. For example, an individual in a deficiency study may not actually exhibit a deficient condition. Any given deficient reading may exhibit excess tendencies, and vice versa. Note that it is important to distinguish the response, \(Y_i\), and the study identifier, \(x_{i2}\). Assuming \(\beta_1 < 0, \beta_3 > 0, \) and \(\beta_1 + \beta_3 > 0\), a display of \(P_E(Y_i = 1)\) and \(P_D(Y_i = 1)\) versus \(x_{i1}\) would appear as in Figure 1.
Figure 1: The components of the JMED: Excess and Deficiency curves.

4.2 Investigating the Point of Intersection

Interest lies in finding a single optimal intake level of Cu where there is a small risk of a severe health effect caused by an excess or deficient condition. To minimize the probability of severity from an excess condition, one would suggest a very small Cu dose. Likewise, to minimize the probability of severity from a deficient condition, one would suggest a very large Cu dose. Essentially, minimizing the likelihood of severity from an excess condition translates into maximizing the chances of severity from a deficient condition, and vice versa.

The crossover(intersection) point between the excess and deficiency curves provides a compromise where the probability of severity from excess exposure is equivalent to the probability of severity from deficiency exposure. Assuming an individual has no apriori knowledge of whether they may be at risk of an excess or deficiency condition, the exposure concentration at the
crossover point will protect an individual from an ‘unknown’ excess or deficient condition. This point will be referred to as the \textit{equiprobable crossover point} (EPCP).

The EPCP has a closed form solution since the intersection point between the two logistic regression equations for excess and deficiency can be solved algebraically. Setting (4.3) equal to (4.2) and solving for $x_{i1}$ will provide an expression for the EPCP.

\[
P_E(Y_i = 1) = P_D(Y_i = 1) \\
\frac{\exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}^*]}{1 + \exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}^*]} = \frac{\exp(\beta_0 + \beta_1 x_{i1}^*)}{1 + \exp(\beta_0 + \beta_1 x_{i1}^*)}
\]

Cross-multiplying, the expression becomes:

\[
\exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}^*] + \exp[(2\beta_0 + \beta_2) + (2\beta_1 + \beta_3)x_{i1}^*] = \exp(\beta_0 + \beta_1 x_{i1}^*) + \exp[(2\beta_0 + \beta_2) + (2\beta_1 + \beta_3)x_{i1}^*]
\]

Moving everything to the LHS:

\[
\exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}^*] - \exp(\beta_0 + \beta_1 x_{i1}^*) = 0 \\
\exp(\beta_0 + \beta_1 x_{i1}^*) [\exp(\beta_2 + \beta_3 x_{i1}^*) - 1] = 0
\]
Then,
\[
\exp(\beta_0 + \beta_1 x_{i1}^*) = 0
\]

or

\[
\exp(\beta_2 + \beta_3 x_{i1}^*) - 1 = 0
\]

Case 1

\[
\exp(\beta_0 + \beta_1 x_{i1}^*) = 0 \Rightarrow \log(\exp(\beta_0 + \beta_1 x_{i1}^*)) = \log(0)
\]

Since \(\log(0) \Rightarrow -\infty\) so there is no solution from this factor.

Case 2

\[
\exp(\beta_2 + \beta_3 x_{i1}^*) - 1 = 0 \Rightarrow \exp(\beta_2 + \beta_3 x_{i1}^*) = 1
\]

\[
\log(\exp(\beta_2 + \beta_3 x_{i1}^*)) = \log(1) \Rightarrow \beta_2 + \beta_3 x_{i1}^* = 0
\]

Solving for \(x_{i1}^*\) :

\[
x_{i1}^* = -\frac{\beta_2}{\beta_3}
\]

The estimator for \(x_{i1}^*\) is

\[
\hat{x}_{i1}^* = -\frac{\hat{\beta}_2}{\hat{\beta}_3}
\]

Therefore,
\[
\hat{\text{EPCP}} = \hat{x}_{i1}^* = -\frac{\hat{\beta}_2}{\hat{\beta}_3}
\]

39
4.3 Properties of the Estimator for EPCP

Using first and second order Taylor Series expansions of $-\frac{\beta_2}{\beta_3}$ about $\hat{\beta}_2 = \beta_2$ and $\hat{\beta}_3 = \beta_3$, approximate expressions for the bias and variance of the estimator, $-\frac{\beta_2}{\beta_3}$, will be derived.

Consider the general case of deriving Taylor Series expansion of $-\frac{\beta_2}{\beta_3}$ about $\hat{\beta}_i = \beta_i$ and $\hat{\beta}_j = \beta_j$.

Denote $f = f(\beta) = f(\beta_i, \beta_j) = -\frac{\beta_i}{\beta_j}$. Then $\hat{f} = f(\hat{\beta}) = f(\hat{\beta}_i, \hat{\beta}_j) = -\frac{\hat{\beta}_i}{\beta_j}$.

The first derivatives are:

$$ \frac{\partial f}{\partial \hat{\beta}_i} = -\frac{1}{\beta_j}, \quad \frac{\partial f}{\partial \hat{\beta}_j} = \frac{\beta_i}{\beta_j^2} $$

while the second derivatives are:

$$ \frac{\partial^2 f}{\partial \hat{\beta}_i^2} = 0, \quad \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_j} = -\frac{2\beta_i}{\beta_j^2}, \quad \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_j} = \frac{1}{\beta_j^2} $$

Thus, a first order Taylor Series expansion of $f(\hat{\beta}) = -\frac{\beta_2}{\beta_3}$ is given by:

$$ f^{(1)}(\hat{\beta}) \approx -\frac{\beta_i}{\beta_j} - \frac{1}{\beta_j} (\hat{\beta}_i - \beta_i) + \frac{\beta_i}{\beta_j^2} (\hat{\beta}_j - \beta_j) \quad (4.5) $$
while a second order Taylor Series expansion is:

\[
\begin{align*}
  f^{(2)}(\hat{\beta}) & \approx -\frac{\beta_i}{\beta_j} - \frac{1}{\beta_j}(\hat{\beta}_i - \beta_i) - 2\frac{\beta_i}{\beta_j^2}(\hat{\beta}_j - \beta_j) - \frac{\beta_i}{\beta_j^3}(\hat{\beta}_j - \beta_j)^2 + \frac{1}{\beta_j^2}(\hat{\beta}_j - \beta_j)(\hat{\beta}_i - \beta_i) \\
\end{align*}
\]  

(4.6)

Using the latter expansion, the bias may be approximated as:

\[
\text{Bias} \left[ f(\hat{\beta}) \right] \approx E \left[ f^{(2)}(\hat{\beta}) \right] - f(\beta) \\
\approx -\frac{\beta_i}{\beta_j} - \frac{1}{\beta_j} E \left[ (\hat{\beta}_j - \beta_j)^2 \right] + \frac{1}{\beta_j^2} E \left[ (\hat{\beta}_j - \beta_j)(\hat{\beta}_i - \beta_i) \right] \\
\Rightarrow \text{Bias} \left[ f(\hat{\beta}) \right] \approx -\frac{\beta_i}{\beta_j^3} \text{Var}(\hat{\beta}_j) + \frac{1}{\beta_j^2} \text{cov}(\hat{\beta}_i, \hat{\beta}_j) 
\]

(4.7)

The first order Taylor series expansion can be employed to propose an approximate expression for \( \text{Var} \left[ \frac{\hat{\beta}_i}{\hat{\beta}_j} \right] \). Specifically,

\[
\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \frac{1}{\beta_j^2} \text{Var}(\hat{\beta}_i) + \frac{\beta_i^2}{\beta_j^4} \text{Var}(\hat{\beta}_j) - 2\frac{\beta_i}{\beta_j^3} \text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \hat{C}^T \hat{\Sigma}_{\hat{\beta}_{i,j}} \hat{C} 
\]

(4.8)

which can be estimated by

\[
\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \frac{1}{\beta_j^2} \text{Var}(\hat{\beta}_i) + \frac{\hat{\beta}_i^2}{\beta_j^4} \text{Var}(\hat{\beta}_j) - 2\frac{\hat{\beta}_i}{\beta_j^3} \text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \hat{C}^T \hat{\Sigma}_{\hat{\beta}_{i,j}} \hat{C} 
\]

(4.9)

where \( \hat{\Sigma}_{\hat{\beta}_{i,j}} \) is a reduced estimated asymptotic covariance matrix comprised of the rows and columns of \( \hat{\Sigma}_\beta \) associated with \( \hat{\beta}_i \) and \( \hat{\beta}_j \), and \( \hat{C} \) is a \( p \)-dimensional vector containing the coefficients \( -\frac{1}{\beta_j} \) and \( \frac{\hat{\beta}_i}{\beta_j^2} \) in equation (4.5) evaluated at \( \beta = \hat{\beta} \), and with the remaining components all zero. The
coefficients are placed in $\hat{C}$ based on the relative positions of $\beta_i$ and $\beta_j$ in the logistic regression model. For example, if $i = 2$ and $j = 3$ in equation (4.5)

$$\hat{C}^T = \left[ 0, 0, -1, \hat{\beta}_2, \frac{\hat{\beta}_2}{\hat{\beta}_3} \right].$$

If the bias in the estimator for EPCP is disregarded, (4.9) may be used to propose a 100(1 - $\alpha$)% large sample confidence interval for $f(\beta) = \frac{\beta_i}{\beta_j}$, namely

$$-\frac{\hat{\beta}_i}{\hat{\beta}_j} \pm z_{1-\frac{\alpha}{2}} \text{ASE} \left[ -\frac{\hat{\beta}_i}{\hat{\beta}_j} \right] (4.10)$$

where $\text{ASE} \left[ -\frac{\hat{\beta}_i}{\hat{\beta}_j} \right] = \sqrt{\text{Var} \left[ f^{(1)}(\hat{\beta}) \right]}$.

In Chapter 5, we outline the methodology employed to obtain $B=1000$ bootstrap estimates of $\hat{\text{EPCP}}^{(s)(b)}$ generated from $S=1000$ simulated data sets. Another option is to construct a 95% bootstrap confidence interval based on the empirical distribution of $\hat{\text{EPCP}}^{(s)(b)}$. A 95% bootstrap confidence interval for EPCP is

$$\left( \hat{\text{EPCP}}^{(s)(b)}_{[0.025]}, \hat{\text{EPCP}}^{(s)(b)}_{[0.975]} \right) (4.11)$$

where $\hat{\text{EPCP}}^{(s)(b)}_{[0.025]}$ and $\hat{\text{EPCP}}^{(s)(b)}_{[0.975]}$ represent the 2.5th and 97.5th quantiles of the empirical distribution of the 1000 values for $\hat{\text{EPCP}}^{(s)(b)}$. 

42
4.4 The Odds of Severity at the EPCP

In addition to inference for the EPCP, of interest may also be the odds of experiencing a severe health effect at the EPCP. Consider the logit function evaluated at the EPCP:

\[
\text{logit} [\pi (\text{EPCP})] = \log \left[ \frac{\pi (\text{EPCP})}{1 - \pi (\text{EPCP})} \right]
\]

This represents an expression for the log odds of experiencing a severe health effect at the EPCP. Note that the expression for the log odds evaluated at the EPCP will be identical for the excess and deficiency cases, since this is the point for log concentration where the two curves intersect.

For simplicity, consider the case of a deficiency reading. When \( x_{i2} = 0 \):

\[
\log \left[ \frac{\pi (\text{EPCP})}{1 - \pi (\text{EPCP})} \right] = \beta_0 + \beta_1 \left( -\frac{\beta_2}{\beta_3} \right) = \beta_0 - \frac{\beta_1 \beta_2}{\beta_3} \quad (4.12)
\]

Using the same approach employed to develop a confidence interval for the EPCP, we begin by considering the general case where \( \hat{\beta}_i = \frac{\hat{\beta}_i}{\hat{\beta}_i} \) is an estimator for the log odds of interest, and utilize the same Taylor Series expansion technique to establish approximate expressions for the bias and variance of this estimator.

Denote \( f = f(\beta) = f(\beta_i, \beta_j, \beta_k, \beta_l) = \beta_i - \frac{\beta_j \beta_k}{\beta_i} \).

Then \( \hat{f} = f(\hat{\beta}) = f(\hat{\beta}_i, \hat{\beta}_j, \hat{\beta}_k, \hat{\beta}_l) = \hat{\beta}_i - \frac{\hat{\beta}_j \hat{\beta}_k}{\hat{\beta}_i} \).

The first and second order Taylor Series expansions of \( f(\hat{\beta}) \) about \( \hat{\beta}_i = \)
\( \beta_i, \hat{\beta}_j = \beta_j, \hat{\beta}_k = \beta_k, \hat{\beta}_l = \beta_l \) make use of the following first and second derivatives:

\[
\begin{align*}
\frac{\partial f}{\partial \hat{\beta}_i} &= 1, \quad \frac{\partial f}{\partial \hat{\beta}_j} = -\frac{\beta_k}{\beta_l}, \quad \frac{\partial f}{\partial \hat{\beta}_k} = -\frac{\beta_j}{\beta_l}, \quad \frac{\partial f}{\partial \hat{\beta}_l} = -\frac{\beta_j \beta_k}{\beta_l^2} \\
\frac{\partial^2 f}{\partial \hat{\beta}_i^2} &= \frac{\partial^2 f}{\partial \hat{\beta}_j^2} = \frac{\partial^2 f}{\partial \hat{\beta}_k^2} = \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_k} = \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_j} = \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_l} = 0 \\
\frac{\partial^2 f}{\partial \hat{\beta}_j \partial \hat{\beta}_k} &= -2\frac{\beta_j \beta_k}{\beta_l^2}, \quad \frac{\partial^2 f}{\partial \hat{\beta}_j \partial \hat{\beta}_k} = -\frac{1}{\beta_l}, \quad \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_k} = \frac{\beta_k}{\beta_l^2}, \quad \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_l} = \frac{\beta_j}{\beta_l^2}
\end{align*}
\]

giving

\[
f^{(1)}(\hat{\beta}) \approx \beta_i - \frac{\beta_j \beta_k}{\beta_l} + \left( \hat{\beta}_i - \beta_i \right) - \frac{\beta_k}{\beta_l} \left( \hat{\beta}_j - \beta_j \right) - \frac{\beta_j}{\beta_l} \left( \hat{\beta}_k - \beta_k \right) + \frac{\beta_j \beta_k}{\beta_l^2} \left( \hat{\beta}_l - \beta_l \right)
\]

(4.13)

and

\[
f^{(2)}(\hat{\beta}) \approx f^{(1)}(\hat{\beta}) - 2\frac{\beta_j \beta_k}{\beta_l^2} \left( \hat{\beta}_i - \beta_i \right)^2 - 2\frac{1}{\beta_l} \left( \hat{\beta}_j - \beta_j \right) \left( \hat{\beta}_k - \beta_k \right)
\]

\[
+ 2\frac{\beta_j}{\beta_l^2} \left( \hat{\beta}_i - \beta_i \right) \left( \hat{\beta}_j - \beta_j \right) + 2\frac{\beta_j}{\beta_l^2} \left( \hat{\beta}_j - \beta_j \right) \left( \hat{\beta}_k - \beta_k \right) + 2\frac{\beta_j}{\beta_l^2} \left( \hat{\beta}_j - \beta_j \right) \left( \hat{\beta}_l - \beta_l \right)
\]

(4.14)

Therefore,

\[
\text{Bias} \left[ f(\hat{\beta}) \right] \approx -2\frac{\beta_j \beta_k}{\beta_l^3} \text{Var} \left( \hat{\beta}_i \right) - 2\frac{1}{\beta_l} \text{cov} \left( \hat{\beta}_j, \hat{\beta}_k \right) + 2\frac{\beta_k}{\beta_l^2} \text{cov} \left( \hat{\beta}_j, \hat{\beta}_l \right) + 2\frac{\beta_j}{\beta_l^2} \text{cov} \left( \hat{\beta}_k, \hat{\beta}_l \right)
\]

(4.15)
and

$$\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \Sigma_\beta \hat{C} \quad (4.16)$$

which is estimated by

$$\hat{\text{Var}} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \hat{\Sigma}_\beta \hat{C} \quad (4.17)$$

where $\hat{\Sigma}_\beta$ is the estimated asymptotic covariance matrix and $\hat{C}$ is a $p$-dimensional vector containing the coefficients $1, -\hat{\beta}_k / \hat{\beta}_l, -\hat{\beta}_j / \hat{\beta}_l$, and $\hat{\beta}_1 \hat{\beta}_k / \hat{\beta}_l^2$ in equation (4.13) evaluated at $\beta = \hat{\beta}$. The coefficients are placed in $\hat{C}$ based on the relative positions of $\beta_i, \beta_j, \beta_k$ and $\beta_l$ in the logistic regression model. For example, for the model defined in equation (4.1), then

$$\hat{C}^T = \left[ 1, -\hat{\beta}_2 / \hat{\beta}_3, -\hat{\beta}_1 / \hat{\beta}_3, \hat{\beta}_1 \hat{\beta}_2 / \hat{\beta}_3^2 \right].$$

As before, ignoring the bias in the estimator, a 100 $(1 - \alpha)$ % large sample confidence interval for the log odds at the EPCP is:

$$\left( \hat{\beta}_i - \frac{\hat{\beta}_j \hat{\beta}_k}{\hat{\beta}_l} \right) \pm z_{1-\frac{\alpha}{2}} \text{ASE} \left[ \hat{\beta}_i - \frac{\hat{\beta}_j \hat{\beta}_k}{\hat{\beta}_l} \right] \quad (4.18)$$

where

$$\text{ASE} \left[ \hat{\beta}_i - \frac{\hat{\beta}_j \hat{\beta}_k}{\hat{\beta}_l} \right] = \sqrt{\text{Var} \left[ f^{(1)}(\hat{\beta}) \right]}.$$

The analogous confidence interval for the odds is obtained by exponentiating the limits of (4.18).
4.5 Comparing the Odds of Excess to Deficiency

Another important consideration in the context of the application considered here is the ability to assess, at a given concentration level, the relative likelihood that a condition of severity is attributable to an excess condition versus a deficient one. In this section, we propose additional inference techniques for estimating the odds ratio of an excess condition to a deficient one at any specified log_{10} concentration level.

Note that the probability of severity associated with a log_{10} concentration equal to \( X \) is:

\[
P(Y = 1 | X) = \pi(X) = \frac{\exp (\beta_0 + \beta_2 + (\beta_1 + \beta_3) X)}{1 + \exp (\beta_0 + \beta_2 + (\beta_1 + \beta_3) X)}
\] (4.19)

Therefore, the log odds of deficiency at a log_{10} concentration equal to \( X \) are

\[
\log \left( \frac{\pi(X)}{1 - \pi(X)} \right) = \beta_0 + \beta_1 X
\] (4.20)

while the analogous log odds for an excess condition are

\[
\log \left( \frac{\pi(X)}{1 - \pi(X)} \right) = \beta_0 + \beta_2 + (\beta_1 + \beta_3) X
\] (4.21)

Combining (4.20) and (4.21), the log odds ratio of an excess to deficiency condition at a log concentration of \( X \) is:

\[
\log [OR] = \beta_2 + \beta_3 X
\] (4.22)
A 100 \((1 - \alpha)\) % large sample confidence interval for the log odds ratio, \(\beta_2 + \beta_3X\), is:

\[
(\hat{\beta}_2 + \hat{\beta}_3X) \pm z_{1-\frac{\alpha}{2}} \text{AŠE} \left[ \hat{\beta}_2 + \hat{\beta}_3X \right]
\]  

(4.23)

where \(\text{AŠE} \left[ \hat{\beta}_2 + \hat{\beta}_3X \right] = \text{Var} \left[ \hat{\beta}_2 \right] + X^2 \text{Var} \left[ \hat{\beta}_3 \right] + 2X \text{cov}(\hat{\beta}_2, \hat{\beta}_3).\)

Exponentiating the limits of (4.23) yields the analogous confidence interval for the odds ratio.

4.6 Stratification by Species

A natural extension of the JMED model is the addition of variables to account for interspecies differences, thereby creating opportunities for species-related analyses. In what follows below, we develop methods for comparing these different species; motivation lies in the examples we present in Chapters 6 and 7 where data is collected on mice, rats, and humans. This requires the addition of indicator variables to the JMED model considered earlier.

As before, we let

\[ x_{i1} = \log_{10} \text{ concentration of the } i^{th} \text{ observation} \]
Now define
\[
x_{i2} = \begin{cases} 
1, & \text{if excess} \\
0, & \text{if deficiency}
\end{cases}
\]

Now define
\[
x_{i3} = \begin{cases} 
1, & \text{if mouse} \\
0, & \text{otherwise}
\end{cases}
\]
\[
x_{i4} = \begin{cases} 
1, & \text{if rat} \\
0, & \text{otherwise}
\end{cases}
\]

All possible two-way interaction terms will also be included in the JMED model. This representation facilitates unique excess and deficiency models for humans, rats, and mice.

The probability of severity may be expressed as:

\[
P(Y_i = 1) = \frac{\exp \left( \sum_{k=0}^{9} x_{ik} \beta_k \right)}{1 + \exp \left( \sum_{k=0}^{9} x_{ik} \beta_k \right)}
\]

where
\[
\sum_{k=0}^{9} x_{ik} \beta_k = \beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i1} x_{i2} + \beta_6 x_{i1} x_{i3} + \beta_7 x_{i1} x_{i4} + \beta_8 x_{i2} x_{i3} + \beta_9 x_{i2} x_{i4}
\]

in which \(x_{i0} = 1, x_{i5} = x_{i1} x_{i2}, x_{i6} = x_{i1} x_{i3}, x_{i7} = x_{i1} x_{i4}, x_{i8} = x_{i2} x_{i3},\)

and \(x_{i9} = x_{i2} x_{i4}\)
4.6.1 Humans

Setting $x_{i3} = x_{i4} = 0$ and $x_{i2} = 0$, the deficiency model for humans is:

$$P_D(Y_i = 1) = \frac{\exp (\beta_0 + \beta_1 x_{i1})}{1 + \exp (\beta_0 + \beta_1 x_{i1})}$$

Setting $x_{i3} = x_{i4} = 0$, and $x_{i2} = 1$, the excess model for humans is:

$$P_E(Y_i = 1) = \frac{\exp (\beta_0 + \beta_2 + (\beta_1 + \beta_5) x_{i1})}{1 + \exp (\beta_0 + \beta_2 + (\beta_1 + \beta_5) x_{i1})}$$

To find the EPCP for humans, set $P_D(Y_i = 1) = P_E(Y_i = 1)$ and solve for $x_{i1}$. The estimator of the EPCP for humans is:

$$\hat{EPCP}_{HU} = -\frac{\hat{\beta}_2}{\hat{\beta}_5}$$

4.6.2 Mice

Setting $x_{i3} = 1, x_{i4} = 0$, and $x_{i2} = 0$, the deficiency model for mice is:

$$P_D(Y_i = 1) = \frac{\exp (\beta_0 + \beta_3 + (\beta_1 + \beta_6) x_{i1})}{1 + \exp (\beta_0 + \beta_3 + (\beta_1 + \beta_6) x_{i1})}$$

Setting $x_{i3} = 1, x_{i4} = 0$, and $x_{i2} = 1$, the excess model for mice is:

$$P_E(Y_i = 1) = \frac{\exp ((\beta_0 + \beta_2 + \beta_3 + \beta_8) + (\beta_1 + \beta_5 + \beta_6) x_{i1})}{1 + \exp ((\beta_0 + \beta_2 + \beta_3 + \beta_8) + (\beta_1 + \beta_5 + \beta_6) x_{i1})}$$

To find the EPCP for mice, set $P_D(Y_i = 1) = P_E(Y_i = 1)$ and solve for $x_{i1}$. The estimator of the EPCP for mice is:
\[ \text{EPCP}_{\text{MU}} = -\frac{\hat{\beta}_2 + \hat{\beta}_8}{\hat{\beta}_5} \]

### 4.6.3 Rats

Setting \( x_{i4} = 1, x_{i3} = 0, \) and \( x_{i2} = 0, \) the deficiency model for rats is:

\[ P_D(Y_i = 1) = \frac{\exp (\beta_0 + \beta_4 + (\beta_1 + \beta_7)x_{i1})}{1 + \exp (\beta_0 + \beta_4 + (\beta_1 + \beta_7)x_{i1})} \]

Setting \( x_{i4} = 1, x_{i3} = 0 \) and \( x_{i2} = 1, \) the excess model for rats is:

\[ P_E(Y_i = 1) = \frac{\exp ((\beta_0 + \beta_2 + \beta_4 + \beta_9) + (\beta_1 + \beta_5 + \beta_7)x_{i1})}{1 + \exp ((\beta_0 + \beta_2 + \beta_4 + \beta_9) + (\beta_1 + \beta_5 + \beta_7)x_{i1})} \]

To find the EPCP for rats, set \( P_D(Y_i = 1) = P_E(Y_i = 1) \) and solve for \( x_{i1}. \) The estimator of the EPCP for rats is:

\[ \text{EPCP}_{\text{RT}} = -\frac{\hat{\beta}_2 + \hat{\beta}_8}{\hat{\beta}_5} \]

### 4.7 Properties of the Estimator for EPCP\text{RT}, EPCP\text{MU}, EPCP\text{HU}

Using the first and second order Taylor Series Expansions of \( -\frac{\hat{\beta}_2}{\hat{\beta}_5} \) about \( \hat{\beta}_2 = \beta_2 \) and \( \hat{\beta}_5 = \beta_5, \) it is possible to approximate expressions for the bias and variance of the EPCP estimator for humans. Since the estimator assumes the form \( -\frac{\hat{\beta}_2}{\hat{\beta}_j}, \) the properties described in Section 4.3 apply.

Using the first and second order Taylor Series expansions of \( -\frac{\hat{\beta}_2 + \hat{\beta}_8}{\hat{\beta}_5} \) about \( \hat{\beta}_2 = \beta_2, \) \( \hat{\beta}_8 = \beta_8, \) and \( \hat{\beta}_5 = \beta_5, \) and \( -\frac{\hat{\beta}_2 + \hat{\beta}_9}{\hat{\beta}_5} \) about \( \hat{\beta}_2 = \beta_2, \) \( \hat{\beta}_9 = \beta_9, \) and
\( \hat{\beta}_5 = \beta_5 \), it is possible to approximate expressions for the bias and variance of the EPCP estimator for mice and rats. Consider the general case of deriving a Taylor Series expansion of \( -\frac{\beta_i + \beta_j}{\beta_k} \) about \( \hat{\beta}_i = \beta_i, \hat{\beta}_j = \beta_j, \) and \( \hat{\beta}_k = \beta_k \).

Denote \( f = f(\beta) = f(\beta_i, \beta_j, \beta_k) = -\frac{\beta_i + \beta_j}{\beta_k} \). Then \( \hat{f} = f(\hat{\beta}) = f(\hat{\beta}_i, \hat{\beta}_j, \hat{\beta}_k) = -\frac{\hat{\beta}_i + \hat{\beta}_j}{\hat{\beta}_k} \).

The first and second order Taylor Series expansions of \( f(\hat{\beta}) \) make use of the following first and second derivatives:

\[
\frac{\partial f}{\partial \hat{\beta}_i} = -\frac{1}{\beta_k}, \quad \frac{\partial f}{\partial \hat{\beta}_j} = -\frac{1}{\beta_k}, \quad \frac{\partial f}{\partial \hat{\beta}_k} = \frac{\beta_i + \beta_j}{\beta_k^2}
\]

\[
\frac{\partial^2 f}{\partial \hat{\beta}_i^2} = \frac{\partial^2 f}{\partial \hat{\beta}_j^2} = 0, \quad \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_j} = 0, \quad \frac{\partial^2 f}{\partial \hat{\beta}_i \partial \hat{\beta}_k} = \frac{1}{\beta_k^2}, \quad \frac{\partial^2 f}{\partial \hat{\beta}_j \partial \hat{\beta}_k} = \frac{1}{\beta_k^2}
\]

giving

\[
f^{(1)}(\hat{\beta}) \approx -\frac{\beta_i + \beta_j}{\beta_k} - \frac{1}{\beta_k} (\hat{\beta}_i - \beta_i) - \frac{1}{\beta_k} (\hat{\beta}_j - \beta_j) + \frac{\beta_i + \beta_j}{\beta_k^2} (\hat{\beta}_k - \beta_k) \quad (4.25)
\]

and

\[
f^{(2)}(\hat{\beta}) \approx -\frac{\beta_i + \beta_j}{\beta_k} - \frac{1}{\beta_k} (\hat{\beta}_i - \beta_i) - \frac{1}{\beta_k} (\hat{\beta}_j - \beta_j) + \frac{\beta_i + \beta_j}{\beta_k^2} (\hat{\beta}_k - \beta_k)
- 2\frac{\beta_i + \beta_j}{\beta_k^3} (\hat{\beta}_k - \beta_k)^2 + \frac{1}{\beta_k^2} (\hat{\beta}_i - \beta_i)(\hat{\beta}_k - \beta_k) + \frac{1}{\beta_k^2} (\hat{\beta}_j - \beta_j)(\hat{\beta}_k - \beta_k)
\]

\[
(4.26)
\]
Therefore,

$$\text{Bias} \left[ f(\hat{\beta}) \right] \approx -2 \frac{\hat{\beta}_i + \hat{\beta}_j}{\hat{\beta}_k^3} \text{Var}(\hat{\beta}_k) + \frac{1}{\hat{\beta}_k^2} \text{cov}(\hat{\beta}_j, \hat{\beta}_k) + \frac{1}{\hat{\beta}_k^2} \text{cov}(\hat{\beta}_i, \hat{\beta}_k)$$

(4.27)

and

$$\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \Sigma_{\beta(i,j,k)} \hat{C}$$

(4.28)

which is estimated by

$$\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \hat{\Sigma}_{\beta(i,j,k)} \hat{C}$$

(4.29)

where $\hat{\Sigma}_{\beta(i,j,k)}$ is a reduced estimated asymptotic covariance matrix comprised of the rows and columns of $\Sigma_{\beta}$ associated with $\hat{\beta}_i$, $\hat{\beta}_j$, and $\hat{\beta}_k$, and $\hat{C}$ is a $p$-dimensional vector containing the coefficients $-\frac{1}{\hat{\beta}_k}$, $-\frac{1}{\hat{\beta}_k}$, and $\frac{\hat{\beta}_i + \hat{\beta}_j}{\hat{\beta}_k^2}$ in equation (4.25) evaluated at $\beta = \hat{\beta}$, with the remaining components all zero. The coefficients are placed in $\hat{C}$ based on the relative positions of $\beta_1$, $\beta_2$, and $\beta_3$ in the logistic regression model. For example, if $i = 2$, $j = 8$, and $k = 5$ for the model defined in equation (4.24), then

$$\hat{C}^T = \left[ 0, 0, -\frac{1}{\hat{\beta}_5}, 0, 0, \frac{\hat{\beta}_2 + \hat{\beta}_5}{\hat{\beta}_5^2}, 0, 0, -\frac{1}{\hat{\beta}_5}, 0 \right].$$

As before, ignoring the bias in the estimator, a 100 $(1 - \alpha)$ % large sample confidence interval for $-\frac{\hat{\beta}_i + \hat{\beta}_j}{\hat{\beta}_k}$ is:

$$-\frac{\hat{\beta}_i + \hat{\beta}_j}{\hat{\beta}_k} \pm z_{1-\alpha} \hat{A}\hat{SE} \left[ -\frac{\hat{\beta}_i + \hat{\beta}_j}{\hat{\beta}_k} \right]$$

(4.30)

52
where $\text{ASE} \left[ -\frac{\hat{\beta}_i + \hat{\beta}_j}{\hat{\beta}_k} \right] = \sqrt{\text{Var} \left[ f^{(1)}(\hat{\beta}) \right]}$

The analogous confidence interval for the EPCP$_{MU}$ or EPCP$_{RT}$ is obtained by exponentiating the limits of (4.30).

\section*{4.8 The Odds of Severity at EPCP$_{HU}$, EPCP$_{MU}$, and EPCP$_{RT}$}

For each species, interest may lie in the odds of experiencing a severe health effect at a log concentration level equal to the EPCP.

\subsection*{4.8.1 Humans}

Consider the logit function evaluated at EPCP$_{HU}$:

$$\text{logit} \left[ \pi (\text{EPCP}_{HU}) \right] = \beta_0 - \beta_1 \left( \frac{\beta_2}{\beta_5} \right) + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} - \beta_5 \left( \frac{\beta_2}{\beta_5} \right) x_{i2}$$

$$- \beta_6 \left( \frac{\beta_2}{\beta_5} \right) x_{i3} - \beta_7 \left( \frac{\beta_2}{\beta_5} \right) x_{i4} + \beta_8 x_{i2} x_{i3} + \beta_9 x_{i2} x_{i4}$$

As in the case of the EPCP for all species, the expression for log odds evaluated at the EPCP will be identical for the excess and deficiency cases. For simplicity, consider the case of a deficiency reading.

When $x_{i2} = 0$ and $x_{i3} = x_{i4} = 0$:

$$\text{log} \left[ \frac{\pi (\text{EPCP}_{HU})}{1 - \pi (\text{EPCP}_{HU})} \right] = \beta_0 + \beta_1 \left( -\frac{\beta_2}{\beta_5} \right) = \beta_0 - \frac{\beta_1 \beta_2}{\beta_5}$$

Using the same approach employed to develop a confidence interval for the log odds at the EPCP for all species, we begin by considering $\hat{\beta}_0 - \frac{\hat{\beta}_1 \hat{\beta}_2}{\hat{\beta}_5}$.
as an estimator for the log odds of interest. Since this estimator assumes the form \( \hat{\beta}_i - \frac{\hat{\beta}_j \hat{\beta}_k}{\hat{\beta}_l} \), we may utilize the properties developed in Section 4.4, with the exception of the variance approximation. The approximation for the variance requires adjustment since there are more covariates in the species stratified JMED model than in the initial JMED model.

An expression for the variance is:

\[
\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \hat{\Sigma}_{\beta(i,j,k,l)} \hat{C}
\]

(4.31)

which may be estimated by

\[
\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \hat{\Sigma}_{\beta(i,j,k,l)} \hat{C}
\]

(4.32)

where \( \hat{\Sigma}_{\beta(i,j,k,l)} \) is a reduced estimated asymptotic covariance matrix comprised of the rows and columns of \( \hat{\Sigma}_\beta \) associated with \( \hat{\beta}_i, \hat{\beta}_j, \hat{\beta}_k, \) and \( \hat{\beta}_l \), and \( \hat{C} \) is a \( p \)-dimensional vector containing the coefficients \( 1, -\frac{\hat{\beta}_k}{\hat{\beta}_l}, -\frac{\hat{\beta}_l}{\hat{\beta}_l}, \) and \( \frac{\hat{\beta}_l \hat{\beta}_k}{\hat{\beta}_l} \) in equation (4.13) evaluated at \( \beta = \hat{\beta} \), with the remaining components all zero. The coefficients are placed in \( \hat{C} \) based on the relative positions of \( \beta_i, \beta_j, \beta_k, \) and \( \beta_l \) in the logistic regression model. For example, for the model defined in (4.24), then

\[
\hat{C} = \left[ 1, -\frac{\hat{\beta}_2}{\hat{\beta}_5}, -\frac{\hat{\beta}_2}{\hat{\beta}_5}, 0, 0, \frac{\hat{\beta}_1 \hat{\beta}_2}{\hat{\beta}_5}, 0, 0, 0, 0 \right].
\]

(4.33)

A 100 \( (1 - \alpha) \) % large sample confidence interval for the log odds at the EPCP for humans may be constructed using the procedure described in

54
4.8.2 Mice and Rats

As we have seen for EPCP_{HU}, the expression for the log odds evaluated at the EPCP_{MU} and EPCP_{RT} will be identical for the excess and deficiency cases.

The log odds at the EPCP for mice is:

\[
\log \left[ \frac{\pi(\text{EPCP}_{MU})}{1 - \pi(\text{EPCP}_{MU})} \right] = \beta_0 + \beta_3 + (\beta_1 + \beta_6) \left( -\frac{(\beta_2 + \beta_8)}{\beta_5} \right) \\
= \beta_0 + \beta_3 - (\beta_1 + \beta_6) \left( \frac{\beta_2 + \beta_8}{\beta_5} \right) \tag{4.34}
\]

The log odds at the EPCP for rats is:

\[
\log \left[ \frac{\pi(\text{EPCP}_{RT})}{1 - \pi(\text{EPCP}_{RT})} \right] = \beta_0 + \beta_4 + (\beta_1 + \beta_7) \left( -\frac{(\beta_2 + \beta_9)}{\beta_5} \right) \\
= \beta_0 + \beta_4 - (\beta_1 + \beta_7) \left( \frac{\beta_2 + \beta_9}{\beta_5} \right) \tag{4.35}
\]

We will consider the general case that is applicable to the log odds at EPCP_{MU} and EPCP_{RT}.

Denote \( f = f(\beta) = f(\beta_0, \beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9) = \beta_0 + \beta_i - (\beta_1 + \beta_j) \left( \frac{\beta_2 + \beta_k}{\beta_5} \right) \)

where \( i = \begin{bmatrix} 3 \\ 6 \\ 8 \end{bmatrix} \) for mice and \( j = \begin{bmatrix} 4 \\ 7 \\ 9 \end{bmatrix} \) for rats.
\[ f = f(\hat{\beta}) = f(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5) = \hat{\beta}_0 + (\hat{\beta}_1 + \hat{\beta}_j) \left( \frac{\hat{\beta}_2 + \hat{\beta}_5}{\hat{\beta}_5} \right). \]

The first and second order Taylor Series expansions of \( f(\hat{\beta}) \) about \( \hat{\beta}_0 = \beta_0, \hat{\beta}_1 = \beta_1, \hat{\beta}_2 = \beta_2, \hat{\beta}_3 = \beta_3, \hat{\beta}_4 = \beta_4, \) and \( \hat{\beta}_5 = \beta_5 \) make use of the following first and second derivatives:

\[
\frac{\partial f}{\partial \hat{\beta}_0} = 1, \quad \frac{\partial f}{\partial \hat{\beta}_1} = \frac{\partial f}{\partial \hat{\beta}_j} = -\frac{\beta_2 + \beta_5}{\beta_5}, \quad \frac{\partial f}{\partial \hat{\beta}_2} = -\frac{\beta_1 + \beta_j}{\beta_5},
\]

\[
\frac{\partial f}{\partial \hat{\beta}_5} = \frac{(\beta_1 + \beta_j)(\beta_2 + \beta_k)}{\beta_5^2}.
\]

\[
\frac{\partial^2 f}{\partial \hat{\beta}_b^2} = 0 \quad \text{for} \quad b = 0, 1, 2, 3, 6, 8 \quad \text{if mice and} \quad b = 0, 1, 2, 4, 7, 9 \quad \text{if rats}
\]

\[
\frac{\partial^2 f}{\partial \hat{\beta}_b \partial \hat{\beta}_0} = 0 \quad \text{for} \quad b = 1, 2, 3, 5, 6, 8 \quad \text{if mice and} \quad b = 1, 2, 4, 5, 7, 9 \quad \text{if rats}.
\]

\[
\frac{\partial^2 f}{\partial \hat{\beta}_1 \partial \hat{\beta}_2} = \frac{\beta_2 + \beta_k}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \hat{\beta}_1 \partial \hat{\beta}_j} = \frac{\beta_2 + \beta_k}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \hat{\beta}_1 \partial \hat{\beta}_5} = 0, \quad \frac{\partial^2 f}{\partial \hat{\beta}_2 \partial \hat{\beta}_5} = \frac{2(\beta_1 + \beta_j)}{\beta_5^2},
\]

\[
\frac{\partial^2 f}{\partial \hat{\beta}_2 \partial \hat{\beta}_j} = -\frac{1}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \hat{\beta}_2 \partial \hat{\beta}_5} = 0, \quad \frac{\partial^2 f}{\partial \hat{\beta}_5 \partial \hat{\beta}_j} = \frac{\beta_2 + \beta_k}{\beta_5^2},
\]

\[
\frac{\partial^2 f}{\partial \hat{\beta}_3 \partial \hat{\beta}_5} = 0, \quad \text{for} \quad b = 0, 1, 2, 5, 6, 8 \quad \text{if mice and} \quad b = 0, 1, 2, 5, 7, 9 \quad \text{if rats}.
\]
\[ \frac{\partial^2 f}{\partial \beta \partial \beta_k} = \frac{\beta_1 + \beta_j}{\beta_5}, \quad \frac{\partial^2 f}{\partial \beta_i \partial \beta_k} = -\frac{1}{\beta_5} \]

giving

\[ f^{(1)}(\hat{\beta}) \approx \beta_0 + \beta_i - (\beta_1 + \beta_j) \left( \frac{\beta_2 + \beta_k}{\beta_5} \right) + \left( \hat{\beta}_0 - \beta_0 \right) - \frac{(\beta_2 + \beta_k)}{\beta_5} (\hat{\beta}_1 - \beta_1) \]
\[ - \frac{\beta_1 + \beta_j}{\beta_5} (\hat{\beta}_2 - \beta_2) + \left( \hat{\beta}_1 - \beta_i \right) \frac{(\beta_1 + \beta_j)(\beta_2 + \beta_k)}{\beta_5^2} (\hat{\beta}_5 - \beta_5) \]
\[ - \frac{(\beta_2 + \beta_k)}{\beta_5} (\hat{\beta}_j - \beta_j) - \frac{(\beta_1 + \beta_j)}{\beta_5} (\hat{\beta}_k - \beta_k) \]
\[ (4.36) \]

and

\[ f^{(2)}(\hat{\beta}) \approx f^{(1)}(\hat{\beta}) - 2 \frac{(\beta_1 + \beta_j)(\beta_2 + \beta_k)}{\beta_5^3} (\hat{\beta}_5 - \beta_5)^2 - \frac{1}{\beta_5} (\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) \]
\[ - \frac{1}{\beta_5} (\hat{\beta}_1 - \beta_1)(\hat{\beta}_j - \beta_j) - \frac{1}{\beta_5} (\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k) + \frac{(\beta_2 + \beta_k)}{\beta_5^2} (\hat{\beta}_1 - \beta_1)(\hat{\beta}_5 - \beta_5) \]
\[ + 2 \frac{(\beta_1 + \beta_j)}{\beta_5^2} (\hat{\beta}_2 - \beta_2)(\hat{\beta}_5 - \beta_5) - \frac{1}{\beta_5} (\hat{\beta}_2 - \beta_2)(\hat{\beta}_j - \beta_j) \]
\[ + \frac{(\beta_2 + \beta_k)}{\beta_5} (\hat{\beta}_5 - \beta_5)(\hat{\beta}_j - \beta_j) + \frac{(\beta_1 + \beta_j)}{\beta_5^2} (\hat{\beta}_5 - \beta_5)(\hat{\beta}_k - \beta_k) \]
\[ - \frac{1}{\beta_5} (\hat{\beta}_j - \beta_j)(\hat{\beta}_k - \beta_k) \]
\[ (4.37) \]

Therefore,

\[ \text{Bias} \left[ f(\hat{\beta}) \right] \approx -2 \frac{(\beta_1 + \beta_j)(\beta_2 + \beta_k)}{\beta_5^3} \text{Var}(\hat{\beta}_5) - \frac{1}{\beta_5} \text{cov}(\hat{\beta}_1, \hat{\beta}_2) - \frac{1}{\beta_5} \text{cov}(\hat{\beta}_1, \hat{\beta}_j) - \frac{1}{\beta_5} \text{cov}(\hat{\beta}_1, \hat{\beta}_k) \]
\[ + \frac{(\beta_2 + \beta_k)}{\beta_5} \text{cov}(\hat{\beta}_1, \hat{\beta}_5) + 2 \frac{(\beta_1 + \beta_j)}{\beta_5^2} \text{cov}(\hat{\beta}_2, \hat{\beta}_5) - \frac{1}{\beta_5} \text{cov}(\hat{\beta}_2, \hat{\beta}_j) \]
\[ + \frac{(\beta_2 + \beta_k)}{\beta_5^2} \text{cov}(\hat{\beta}_5, \hat{\beta}_j) + \frac{(\beta_1 + \beta_j)}{\beta_5^2} \text{cov}(\hat{\beta}_5, \hat{\beta}_k) - \frac{1}{\beta_5} \text{cov}(\hat{\beta}_j, \hat{\beta}_k) \]
\[ (4.38) \]
and

$$\text{Var} \left[ f^{(1)}(\beta) \right] \approx \hat{C}^T \hat{\Sigma}_{\beta(0,1,2,3,5,6)} \hat{C} \quad (4.39)$$

which may be estimated by

$$\hat{\text{Var}} \left[ f^{(1)}(\beta) \right] \approx \hat{C}^T \hat{\Sigma}_{\beta(0,1,2,3,5,6)} \hat{C} \quad (4.40)$$

where \( \hat{\Sigma}_{\beta(0,1,2,3,5,6)} \) is a reduced estimated asymptotic covariance matrix comprised of the rows and columns of \( \hat{\Sigma}_{\beta} \) associated with \( \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_5, \hat{\beta}_6 \), and \( \hat{\beta}_k \), and \( \hat{C} \) is a \( p \)-dimensional vector containing the coefficients \( 1, -\frac{\hat{\beta}_2 + \hat{\beta}_k}{\hat{\beta}_5}, -\frac{\hat{\beta}_1 + \hat{\beta}_2}{\hat{\beta}_5}, \frac{(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4)}{\hat{\beta}_5}, -\frac{\hat{\beta}_2 + \hat{\beta}_3}{\hat{\beta}_5}, -\frac{\hat{\beta}_1 + \hat{\beta}_4}{\hat{\beta}_5}, 0, 0 \) in equation (4.36) evaluated at \( \beta = \hat{\beta} \), with the remaining components all zero. The coefficients are placed in \( \hat{C} \) based on the relative positions of \( \beta_0, \beta_1, \beta_2, \beta_3, \beta_5, \beta_6 \), and \( \beta_k \) in the logistic regression model. For example, for the model defined in equation (4.24), then

\[
\hat{C}^T = \begin{bmatrix} 1, -\frac{\hat{\beta}_2 + \hat{\beta}_k}{\hat{\beta}_5}, -\frac{\hat{\beta}_1 + \hat{\beta}_2}{\hat{\beta}_5}, 1, 0, \frac{(\hat{\beta}_1 + \hat{\beta}_2)(\hat{\beta}_2 + \hat{\beta}_3)}{\hat{\beta}_5}, -\frac{\hat{\beta}_2 + \hat{\beta}_3}{\hat{\beta}_5}, 0, -\frac{\hat{\beta}_1 + \hat{\beta}_4}{\hat{\beta}_5}, 0 \end{bmatrix}.
\]

If we ignore the bias in the estimator, a 100 \( 1 - \alpha \) % large sample confidence interval for the log odds at the EPCP for mice or rats is

$$\left( \hat{\beta}_0 + \hat{\beta}_i - \frac{(\hat{\beta}_2 + \hat{\beta}_k)(\hat{\beta}_1 + \hat{\beta}_j)}{\hat{\beta}_5} \right) \pm z_{1-\alpha/2} \text{ASE} \left[ \hat{\beta}_0 + \hat{\beta}_i - \frac{(\hat{\beta}_2 + \hat{\beta}_k)(\hat{\beta}_1 + \hat{\beta}_j)}{\hat{\beta}_5} \right] \quad (4.41)$$

where \( \text{ASE} \left[ \hat{\beta}_0 + \hat{\beta}_i - \frac{(\hat{\beta}_2 + \hat{\beta}_k)(\hat{\beta}_1 + \hat{\beta}_j)}{\hat{\beta}_5} \right] = \sqrt{\text{Var} \left[ f^{(1)}(\beta) \right]} \).
Exponentiating the limits of (4.41) yields the analogous confidence interval for the odds at the EPCP.

4.9 Species-to-Species Odds Ratios

We develop odds ratios (OR) that allow us to compare two species at their respective EPCP’s. Using these odds ratios, we can determine which species is more likely to experience a non-normal reading attributable to a log_{10} concentration level equal to the EPCP.

4.9.1 Mice to Humans and Rats to Humans

The odds expressions for mice, rats, and humans experiencing a non-normal reading at a log concentration level equal to the EPCP are

\[
\text{Odds}_{MU} = \frac{\pi_{MU}}{1 - \pi_{MU}} = \frac{\pi \left( \text{EPCP}_{MU} \right)}{1 - \pi \left( \text{EPCP}_{MU} \right)} = \exp \left( \beta_0 + \beta_3 - \frac{(\beta_2 + \beta_8)(\beta_1 + \beta_6)}{\beta_5} \right) \\
\text{(4.42)}
\]

\[
\text{Odds}_{RT} = \frac{\pi_{RT}}{1 - \pi_{RT}} = \frac{\pi \left( \text{EPCP}_{RT} \right)}{1 - \pi \left( \text{EPCP}_{RT} \right)} = \exp \left( \beta_0 + \beta_4 - \frac{(\beta_2 + \beta_9)(\beta_1 + \beta_7)}{\beta_5} \right) \\
\text{(4.43)}
\]

\[
\text{Odds}_{HU} = \frac{\pi_{HU}}{1 - \pi_{HU}} = \frac{\pi \left( \text{EPCP}_{HU} \right)}{1 - \pi \left( \text{EPCP}_{HU} \right)} = \exp \left( \beta_0 - \frac{\beta_1 \beta_2}{\beta_5} \right) \\
\text{(4.44)}
\]
Dividing (4.42) and (4.44), the odds ratio for mice to humans is

\[
\frac{\text{Odds}_{\text{MU}}}{\text{Odds}_{\text{HU}}} = \frac{\exp(\beta_0 + \beta_3 - \frac{(\beta_2 + \beta_4)(\beta_1 + \beta_6)}{\beta_5})}{\exp(\beta_0 - \frac{\beta_1 \beta_2}{\beta_5})} = \exp\left(\beta_3 - \frac{(\beta_2 \beta_6 + \beta_1 \beta_k + \beta_6 \beta_8)}{\beta_5}\right)
\]

(4.45)

Dividing (4.43) and (4.44), the odds ratio for rats to humans is

\[
\frac{\text{Odds}_{\text{RT}}}{\text{Odds}_{\text{HU}}} = \frac{\exp(\beta_0 + \beta_4 - \frac{(\beta_2 + \beta_6)(\beta_1 + \beta_7)}{\beta_5})}{\exp(\beta_0 - \frac{\beta_1 \beta_2}{\beta_5})} = \exp\left(\beta_4 - \frac{(\beta_2 \beta_7 + \beta_1 \beta_6 + \beta_7 \beta_9)}{\beta_5}\right)
\]

(4.46)

We will consider the general case where the estimator for the log odds ratio is given by

\[
\log (\hat{\text{OR}}) = \hat{\beta}_i - (\hat{\beta}_2 \hat{\beta}_j + \hat{\beta}_1 \hat{\beta}_k + \hat{\beta}_j \hat{\beta}_k)\hat{\beta}_5
\]

where

\[
\begin{bmatrix}
i \\
j \\
k
\end{bmatrix} = \begin{bmatrix}3 \\6 \\8 \end{bmatrix} \text{ for mice}
\]

and

\[
\begin{bmatrix}
i \\
j \\
k
\end{bmatrix} = \begin{bmatrix}4 \\7 \\9 \end{bmatrix} \text{ for rats.}
\]

Denote \(f(\beta) = f(\beta_1, \beta_2, \beta_3, \beta_5, \beta_j, \beta_k) = \beta_i - (\beta_2 \beta_j + \beta_1 \beta_k + \beta_j \beta_k)\beta_5\). Then \(f(\hat{\beta}) = f(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_5, \hat{\beta}_j, \hat{\beta}_k) = \hat{\beta}_i - (\hat{\beta}_2 \hat{\beta}_j + \hat{\beta}_1 \hat{\beta}_k + \hat{\beta}_j \hat{\beta}_k)\beta_5\).

The first and second order Taylor series expansions of \(f(\hat{\beta})\) about \(\hat{\beta} = \beta\) make use of the following first and second derivatives

\[
\begin{align*}
\frac{\partial f}{\partial \beta_0} = \frac{\partial f}{\partial \beta_i} = \frac{\partial f}{\partial \beta_j} = \frac{\partial f}{\partial \beta_k} = 0, & \quad \frac{\partial f}{\partial \beta_1} = -\frac{\beta_k}{\beta_5}, \quad \frac{\partial f}{\partial \beta_2} = -\frac{\beta_j}{\beta_5}, \quad \frac{\partial f}{\partial \beta_3} = 1 \\
\frac{\partial f}{\partial \beta_5} = \frac{(\beta_2 \beta_j + \beta_1 \beta_5 + \beta_j \beta_k)}{\beta_5^2}, & \quad \frac{\partial f}{\partial \beta_j} = -\frac{\beta_2 + \beta_k}{\beta_5}, \quad \frac{\partial f}{\partial \beta_k} = -\frac{\beta_1 + \beta_j}{\beta_5}
\end{align*}
\]
\[
\frac{\partial^2 f}{\partial \beta^2 b} = 0 \quad \text{for} \quad b = 0, \ldots, 9; \quad \frac{\partial^2 f}{\partial \beta_b \partial \beta_0} = 0 \quad \text{for} \quad b = 1, \ldots, 9
\]

\[
\frac{\partial^2 f}{\partial \beta_b \partial \beta_1} = 0 \quad \text{if} \quad b = 2, 3, 4, 6, 7, 9; \quad \text{for mice and} \quad b = 2, 3, 4, 6, 7, 8 \quad \text{for rats}
\]

\[
\frac{\partial^2 f}{\partial \beta_b \partial \beta_2} = 0 \quad \text{if} \quad b = 3, 4, 7, 8, 9 \quad \text{for mice and} \quad b = 3, 4, 6, 8, 9 \quad \text{for rats}
\]

\[
\frac{\partial^2 f}{\partial \beta_b \partial \beta_3} = 0 \quad \text{if} \quad b = 4, 5, 6, 7, 8, 9 \quad \text{for mice and} \quad b = 3, 5, 6, 7, 8, 9 \quad \text{for rats}
\]

\[
\frac{\partial^2 f}{\partial \beta_b \partial \beta_i} = 0 \quad \text{for} \quad b = 5, 6, 7, 8, 9
\]

\[
\frac{\partial^2 f}{\partial \beta_b \partial \beta_j} = \frac{\partial^2 f}{\partial \beta_b \partial \beta_j} = 0 \quad \text{if} \quad b = 7, 9 \quad \text{for mice and} \quad b = 6, 8 \quad \text{for rats}
\]

\[
\frac{\partial^2 f}{\partial \beta_2 \partial \beta_5} = \frac{\partial^2 f}{\partial \beta_2 \partial \beta_j} = \frac{-2 \beta_2 \beta_j + \beta_1 \beta_5 + \beta_j \beta_k}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_1 \partial \beta_5} = \frac{\beta_k}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_1 \partial \beta_k} = \frac{1}{\beta_5}
\]

\[
\frac{\partial^2 f}{\partial \beta_2 \partial \beta_5} = \frac{\beta_j}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_2 \partial \beta_j} = \frac{-1}{\beta_5}, \quad \frac{\partial^2 f}{\partial \beta_3 \partial \beta_5} = \frac{(\beta_2 + \beta_k)}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_3 \partial \beta_k} = \frac{-1}{\beta_5}
\]

giving

\[
f^{(1)}(\hat{\beta}) \approx f(\beta) - \frac{\beta_k}{\beta_5}(\hat{\beta}_1 - \beta_1) - \frac{\beta_j}{\beta_5}(\hat{\beta}_2 - \beta_2) + (\hat{\beta}_1 - \beta_1) + \frac{\beta_2 \beta_j + \beta_1 \beta_5 + \beta_j \beta_k}{\beta_5^2}(\hat{\beta}_5 - \beta_5) - \frac{\beta_2 + \beta_k}{\beta_5}(\hat{\beta}_j - \beta_j) - \frac{\beta_1 + \beta_j}{\beta_5}(\hat{\beta}_k - \beta_k)
\]  

(4.47)

and

\[
f^{(2)}(\hat{\beta}) \approx f^{(1)}(\hat{\beta}) - 2\frac{\beta_2 \beta_j + \beta_1 \beta_5 + \beta_j \beta_k}{\beta_5^2}(\hat{\beta}_5 - \beta_5)^2 + \frac{\beta_k}{\beta_5^2}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_5 - \beta_5) - \frac{1}{\beta_5}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k) + \frac{\beta_j}{\beta_5^2}(\hat{\beta}_2 - \beta_2)(\hat{\beta}_5 - \beta_5) - \frac{1}{\beta_5}(\hat{\beta}_2 - \beta_2)(\hat{\beta}_j - \beta_j) + \frac{\beta_1 + \beta_k}{\beta_5^2}(\hat{\beta}_5 - \beta_5)(\hat{\beta}_k - \beta_k) - \frac{1}{\beta_5}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k)
\]

61
Therefore,

\[
\text{Bias}[f(\hat{\beta})] \approx -2\frac{\beta_2\beta_j + \beta_1\beta_5 + \beta_j\beta_k}{\beta_5^2}\text{Var}(\hat{\beta}_5) + \frac{\beta_k}{\beta_5^2}\text{cov}(\hat{\beta}_1, \hat{\beta}_5) - \frac{1}{\beta_5}\text{cov}(\hat{\beta}_1, \hat{\beta}_k) + \frac{\beta_j}{\beta_5^2}\text{cov}(\hat{\beta}_2, \hat{\beta}_5)
- \frac{1}{\beta_5}\text{cov}(\hat{\beta}_2, \hat{\beta}_j) + \frac{\beta_2 + \beta_k}{\beta_5^2}\text{cov}(\hat{\beta}_5, \hat{\beta}_j) + \frac{\beta_1 + \beta_j}{\beta_5^2}\text{cov}(\hat{\beta}_5, \hat{\beta}_k) - \frac{1}{\beta_5}\text{cov}(\hat{\beta}_j, \hat{\beta}_k)
\]

(4.48)

and

\[
\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \hat{\Sigma}_{\beta(1,2,i,5,j,k)} \hat{C}
\]

(4.49)

which is estimated by

\[
\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \hat{\Sigma}_{\beta(1,2,i,5,j,k)} \hat{C}
\]

(4.50)

where \(\hat{\Sigma}_{\beta(1,2,i,5,j,k)}\) is a reduced estimated asymptotic covariance matrix comprised of the rows and columns of \(\hat{\Sigma}_{\beta}\) associated with \(\beta_1, \beta_2, \beta_i, \beta_5, \beta_j, \text{ and } \beta_k\) and \(\hat{C}\) is a \(p\)-dimensional vector containing the coefficients \(1, \frac{\hat{\beta}_k}{\beta_5}, -\frac{\hat{\beta}_j}{\beta_5}, 1, \frac{\hat{\beta}_2 + \hat{\beta}_5}{\beta_5}, \frac{\beta_1 + \hat{\beta}_i}{\beta_5}\), and \(\frac{\beta_1 + \hat{\beta}_i}{\beta_5}\) in equation (4.47) evaluated at \(\beta = \hat{\beta}\), with the remaining components all zero. The coefficients are placed in \(\hat{C}\) based on the relative positions of \(\beta_1, \beta_2, \beta_i, \beta_5, \beta_j, \text{ and } \beta_k\) in the logistic regression model. For example, for the model defined in equation (4.24), then

\[
\hat{C}^T = \begin{bmatrix}
1, \frac{\hat{\beta}_k}{\beta_5}, -\frac{\hat{\beta}_j}{\beta_5}, 1, 0, \frac{\hat{\beta}_2 + \hat{\beta}_5 + \hat{\beta}_j\hat{\beta}_k}{\beta_5^2}, -\frac{\hat{\beta}_2 + \hat{\beta}_k}{\beta_5}, 0, -\frac{\hat{\beta}_1 + \hat{\beta}_i}{\beta_5}, 0
\end{bmatrix}
\]

Ignoring the bias in the estimator, the general form for a 100 \((1 - \alpha)\) % large sample confidence interval for the mice to humans and rats to humans odds
ratio is:

\[
\left( \hat{\beta}_i - \frac{(\hat{\beta}_2 \hat{\beta}_j + \hat{\beta}_1 \hat{\beta}_k + \hat{\beta}_j \hat{\beta}_k)}{\hat{\beta}_5} \right) \pm z_{1-\frac{1}{2}} \text{ASE} \left[ \hat{\beta}_i - \frac{(\hat{\beta}_2 \hat{\beta}_j + \hat{\beta}_1 \hat{\beta}_k + \hat{\beta}_j \hat{\beta}_k)}{\hat{\beta}_5} \right]
\]

where \( \text{ASE} \left[ \hat{\beta}_i - \frac{(\hat{\beta}_2 \hat{\beta}_j + \hat{\beta}_1 \hat{\beta}_k + \hat{\beta}_j \hat{\beta}_k)}{\hat{\beta}_5} \right] = \sqrt{\text{Var} \left[ f(1)(\hat{\beta}) \right]} \)

(4.51)

Exponentiating the limits of (4.51) yields the analogous confidence interval for the odds ratio.

### 4.9.2 Rats to Mice

Dividing (4.43) and (4.42), the odds ratio for rats to mice is:

\[
\text{Odds}_{\text{RT}} \quad \text{Odds}_{\text{MU}} = \exp \left( \beta_0 + \beta_4 - \frac{(\beta_2 + \beta_5)(\beta_4 + \beta_7)}{\beta_5} \right)
\]

\[
= \exp \left( \beta_4 - \beta_3 - \frac{\beta_2 \beta_7 - \beta_1 \beta_9 - \beta_7 \beta_3 + \beta_2 \beta_6 + \beta_4 \beta_8 + \beta_1 \beta_8 + \beta_6 \beta_8}{\beta_5} \right)
\]

(4.52)

We shall use a Taylor series expansion to linearize the log odds ratio.

Denote \( f(\beta) = f(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9) = \beta_4 - \beta_3 - \frac{\beta_2 \beta_7 - \beta_1 \beta_9 - \beta_7 \beta_3 + \beta_2 \beta_6 + \beta_4 \beta_8 + \beta_1 \beta_8 + \beta_6 \beta_8}{\beta_5} \).

Then \( f(\hat{\beta}) = f(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6, \hat{\beta}_7, \hat{\beta}_8, \hat{\beta}_9) = \hat{\beta}_4 - \hat{\beta}_3 - \frac{\beta_2 \beta_7 - \beta_1 \beta_9 - \beta_7 \beta_3 + \beta_2 \beta_6 + \beta_4 \beta_8 + \beta_1 \beta_8 + \beta_6 \beta_8}{\beta_5} \).

As we have seen, the first and second order Taylor series expansions of \( f(\hat{\beta}) \) about \( \hat{\beta} = \beta \) make use of the following first and second derivatives

\[
\frac{\partial f}{\partial \beta_0} = 0, \quad \frac{\partial f}{\partial \hat{\beta}_1} = -\frac{\beta_3 + \beta_8}{\beta_5}, \quad \frac{\partial f}{\partial \hat{\beta}_2} = -\frac{\beta_7 + \beta_6}{\beta_5}, \quad \frac{\partial f}{\partial \hat{\beta}_3} = -1, \quad \frac{\partial f}{\partial \hat{\beta}_4} = 1
\]
\[ \frac{\partial f}{\partial \beta_5} = \frac{\beta_2 \beta_7 + \beta_1 \beta_0 + \beta_7 \beta_0 - \beta_2 \beta_6 - \beta_1 \beta_5 - \beta_6 \beta_8}{\beta_5^2}, \quad \frac{\partial f}{\partial \beta_6} = \frac{\beta_2 + \beta_8}{\beta_5}, \]

\[ \frac{\partial f}{\partial \beta_7} = -\frac{\beta_2 + \beta_5}{\beta_5}, \quad \frac{\partial f}{\partial \beta_8} = \frac{\beta_1 + \beta_5}{\beta_5}, \quad \frac{\partial f}{\partial \beta_9} = -\frac{\beta_1 + \beta_7}{\beta_5}. \]

\[ \frac{\partial^2 f}{\partial \beta_j^2} = 0 \text{ for } j = 0, \ldots, 4; \quad \frac{\partial^2 f}{\partial \beta_j \partial \beta_0} = 0 \text{ for } j = 1, \ldots, 9 \]

\[ \frac{\partial^2 f}{\partial \beta_j \partial \beta_1} = 0 \text{ for } j = 2, 3, 4, 6, 7; \quad \frac{\partial^2 f}{\partial \beta_j \partial \beta_2} = 0 \text{ for } j = 3, 4, 8, 9 \]

\[ \frac{\partial^2 f}{\partial \beta_j \partial \beta_3} = 0 \text{ for } j = 4, 5, 6, 7, 8, 9; \quad \frac{\partial^2 f}{\partial \beta_j \partial \beta_4} = 0 \text{ for } j = 5, 6, 7, 8, 9 \]

\[ \frac{\partial^2 f}{\partial \beta_j \partial \beta_5} = 0 \text{ for } j = 7, 9; \quad \frac{\partial^2 f}{\partial \beta_j \partial \beta_7} = 0 \text{ for } j = 8; \quad \frac{\partial^2 f}{\partial \beta_j \partial \beta_8} = 0 \text{ for } j = 9 \]

\[ \frac{\partial^2 f}{\partial \beta_5^2} = -2 \frac{\beta_2 \beta_7 + \beta_1 \beta_0 + \beta_7 \beta_0 - \beta_2 \beta_6 - \beta_1 \beta_5 - \beta_6 \beta_8}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_6^2} = \frac{\beta_2 + \beta_8}{\beta_5}, \quad \frac{\partial^2 f}{\partial \beta_7^2} = -\frac{\beta_2 + \beta_9}{\beta_5} \]

\[ \frac{\partial^2 f}{\partial \beta_1 \partial \beta_5} = \frac{\beta_9 - \beta_8}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_1 \partial \beta_6} = \frac{\beta_1}{\beta_5}, \quad \frac{\partial^2 f}{\partial \beta_1 \partial \beta_7} = -\frac{\beta_1}{\beta_5}, \quad \frac{\partial^2 f}{\partial \beta_2 \partial \beta_5} = \frac{\beta_7 - \beta_6}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_2 \partial \beta_6} = \frac{\beta_1}{\beta_5}, \quad \frac{\partial^2 f}{\partial \beta_2 \partial \beta_7} = -\frac{\beta_1}{\beta_5}, \quad \frac{\partial^2 f}{\partial \beta_3 \partial \beta_5} = -\frac{\beta_2 + \beta_8}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_3 \partial \beta_6} = -\frac{\beta_1 + \beta_6}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_3 \partial \beta_7} = \frac{\beta_1 + \beta_7}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_4 \partial \beta_5} = -\frac{1}{\beta_5}, \quad \frac{\partial^2 f}{\partial \beta_4 \partial \beta_6} = \frac{\beta_1 + \beta_7}{\beta_5^2}, \quad \frac{\partial^2 f}{\partial \beta_4 \partial \beta_7} = -\frac{1}{\beta_5} \]
giving

\[ f^{(1)}(\hat{\beta}) \approx f(\beta) - \frac{(\beta_0 + \beta_8)}{\beta_5} (\hat{\beta}_1 - \beta_1) - \frac{(\beta_7 + \beta_6)}{\beta_5} (\hat{\beta}_2 - \beta_2) - (\hat{\beta}_3 - \beta_3) + (\hat{\beta}_4 - \beta_4) + \frac{\beta_2 \beta_7 + \beta_1 \beta_9 + \beta_7 \beta_9 - \beta_2 \beta_6 - \beta_1 \beta_8 - \beta_6 \beta_8 (\hat{\beta}_5 - \beta_5)}{\beta_5^2} \]

\[ + \frac{\beta_2 + \beta_8}{\beta_5} (\hat{\beta}_6 - \beta_6) - \frac{\beta_2 + \beta_9}{\beta_5} (\hat{\beta}_7 - \beta_7) + \frac{\beta_1 + \beta_6}{\beta_5} (\hat{\beta}_8 - \beta_8) - \frac{\beta_1 + \beta_7}{\beta_5} (\hat{\beta}_9 - \beta_9) \]

(4.53)

and

\[ f^{(2)}(\hat{\beta}) \approx f^{(1)}(\hat{\beta}) - \frac{2 \beta_2 \beta_7 + \beta_1 \beta_9 + \beta_7 \beta_9 - \beta_2 \beta_6 - \beta_1 \beta_8 - \beta_6 \beta_8 (\hat{\beta}_5 - \beta_5)^2}{\beta_5^2} + \frac{\beta_2 + \beta_8}{\beta_5} (\hat{\beta}_6 - \beta_6)^2 \]

\[ - \frac{\beta_2 + \beta_9}{\beta_5} (\hat{\beta}_7 - \beta_7)^2 + \frac{\beta_1 + \beta_6}{\beta_5} (\hat{\beta}_8 - \beta_8)^2 - \frac{\beta_1 + \beta_7}{\beta_5} (\hat{\beta}_9 - \beta_9)^2 \]

\[ + \frac{\beta_0 - \beta_8}{\beta_5^2} (\hat{\beta}_1 - \beta_1)(\hat{\beta}_5 - \beta_5) + \frac{\beta_1 - \beta_6}{\beta_5^2} (\hat{\beta}_2 - \beta_2)(\hat{\beta}_5 - \beta_5) + \frac{\beta_1}{\beta_5} (\hat{\beta}_2 - \beta_2)(\hat{\beta}_6 - \beta_6) \]

\[ - \frac{\beta_1}{\beta_5} (\hat{\beta}_2 - \beta_2)(\hat{\beta}_7 - \beta_7) + \frac{\beta_1}{\beta_5} (\hat{\beta}_1 - \beta_1)(\hat{\beta}_8 - \beta_8) - \frac{\beta_1}{\beta_5} (\hat{\beta}_1 - \beta_1)(\hat{\beta}_9 - \beta_9) \]

\[ - \frac{\beta_2 + \beta_8}{\beta_5} (\hat{\beta}_5 - \beta_5)(\hat{\beta}_6 - \beta_6) + \frac{\beta_2 + \beta_9}{\beta_5^2} (\hat{\beta}_5 - \beta_5)(\hat{\beta}_7 - \beta_7) - \frac{\beta_1 + \beta_6}{\beta_5^2} (\hat{\beta}_5 - \beta_5)(\hat{\beta}_8 - \beta_8) \]

\[ + \frac{\beta_1 + \beta_7}{\beta_5^2} (\hat{\beta}_5 - \beta_5)(\hat{\beta}_9 - \beta_9) + \frac{1}{\beta_5} (\hat{\beta}_6 - \beta_6)(\hat{\beta}_8 - \beta_8) - \frac{1}{\beta_5} (\hat{\beta}_7 - \beta_7)(\hat{\beta}_9 - \beta_9) \]

(4.54)
Therefore,

\[
\text{Bias} [f(\hat{\beta})] \approx -2 \frac{\beta_2 \beta_7 + \beta_1 \beta_9 + \beta_7 \beta_9 - \beta_2 \beta_6 - \beta_1 \beta_8 - \beta_6 \beta_8 \text{Var}(\hat{\beta}_5) + \frac{\beta_2 + \beta_8}{\beta_5} \text{Var}(\hat{\beta}_6)}{\beta_5^2}
\]

\[
- \frac{\beta_2 + \beta_8}{\beta_5} \text{Var}(\hat{\beta}_7) + \frac{\beta_1 + \beta_6}{\beta_5} \text{Var}(\hat{\beta}_8) - \frac{\beta_1 + \beta_7}{\beta_5} \text{Var}(\hat{\beta}_9) + \frac{\beta_9 - \beta_8}{\beta_5^2} \text{cov}(\hat{\beta}_1, \hat{\beta}_5)
\]

\[
+ \frac{\beta_7 - \beta_6}{\beta_5^2} \text{cov}(\hat{\beta}_2, \hat{\beta}_5) + \frac{\beta_1}{\beta_5} \text{cov}(\hat{\beta}_2, \hat{\beta}_6) - \frac{\beta_1}{\beta_5} \text{cov}(\hat{\beta}_2, \hat{\beta}_7) + \frac{\beta_1}{\beta_5} \text{cov}(\hat{\beta}_1, \hat{\beta}_8)
\]

\[
- \frac{\beta_1}{\beta_5} \text{cov}(\hat{\beta}_1, \hat{\beta}_9) - \frac{\beta_2 + \beta_8}{\beta_5} \text{cov}(\hat{\beta}_5, \hat{\beta}_6) + \frac{\beta_2 + \beta_9}{\beta_5^2} \text{cov}(\hat{\beta}_5, \hat{\beta}_7) - \frac{\beta_1 + \beta_6}{\beta_5^2} \text{cov}(\hat{\beta}_5, \hat{\beta}_8)
\]

\[
+ \frac{\beta_1 + \beta_7}{\beta_5^2} \text{cov}(\hat{\beta}_5, \hat{\beta}_9) + \frac{1}{\beta_5} \text{cov}(\hat{\beta}_6, \hat{\beta}_8) - \frac{1}{\beta_5} \text{cov}(\hat{\beta}_7, \hat{\beta}_9)
\] (4.55)

and

\[
\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \Sigma_{\hat{\beta}_{1,\ldots,9}} \hat{C}
\] (4.56)

which can be estimated by

\[
\text{Var} \left[ f^{(1)}(\hat{\beta}) \right] \approx \hat{C}^T \hat{\Sigma}_{\hat{\beta}_{1,\ldots,9}} \hat{C}
\] (4.57)

where \( \hat{\Sigma}_{\hat{\beta}_{1,\ldots,9}} \) is a reduced estimated asymptotic covariance matrix comprised of the rows and columns of \( \hat{\Sigma}_{\beta} \) associated with \( \hat{\beta}_1, \ldots, \hat{\beta}_9 \), and \( \hat{C} \) is a \( p \)-dimensional vector containing the coefficients \(-\frac{\beta_4 - \beta_5}{\beta_5}, -\frac{\beta_7 - \beta_6}{\beta_5}, -1, 1, \)

\( \frac{(\hat{\beta}_2 \hat{\beta}_7 + \hat{\beta}_1 \hat{\beta}_9 + \hat{\beta}_7 \hat{\beta}_9 - \hat{\beta}_2 \hat{\beta}_6 - \hat{\beta}_1 \hat{\beta}_8 - \hat{\beta}_6 \hat{\beta}_8)}{\beta_5^2}, -\frac{\beta_2 + \beta_8}{\beta_5}, -\frac{\beta_2 + \beta_9}{\beta_5}, -\frac{\beta_1 + \beta_7}{\beta_5} \) in equation (4.53) evaluated at \( \beta = \hat{\beta} \), and with the remaining component zero. The coefficients are placed in \( \hat{C} \) based on the relative positions of \( \beta_1, \ldots, \beta_9 \) in the logistic regression model. For example, for the model defined in equation
\( \hat{C} = \left[ 0, -\frac{\beta_6 - \beta_8}{\beta_5}, -\frac{\beta_7 - \beta_6}{\beta_5}, -1, 1, \left( \frac{\hat{\beta}_2 \hat{\beta}_7 + \hat{\beta}_1 \hat{\beta}_3 + \hat{\beta}_7 \hat{\beta}_6 - \hat{\beta}_2 \hat{\beta}_6 - \hat{\beta}_1 \hat{\beta}_6 - \hat{\beta}_6 \hat{\beta}_6}{\beta_5^2} \right), -\frac{\hat{\beta}_2 + \hat{\beta}_3}{\beta_5}, -\frac{\hat{\beta}_3 + \hat{\beta}_4}{\beta_5}, -\frac{\hat{\beta}_1 + \hat{\beta}_5}{\beta_5}, -\frac{\hat{\beta}_1 + \hat{\beta}_7}{\beta_5} \right]^T \)

Ignoring the bias in the estimator, a 100 \((1 - \alpha)\)% large sample confidence interval for the mice to rats log odds ratio is:

\[
\left( \hat{\beta}_4 - \hat{\beta}_3 - \left( \frac{\hat{\beta}_2 \hat{\beta}_7 + \hat{\beta}_1 \hat{\beta}_3 + \hat{\beta}_7 \hat{\beta}_6 - \hat{\beta}_2 \hat{\beta}_6 - \hat{\beta}_1 \hat{\beta}_6 - \hat{\beta}_6 \hat{\beta}_6}{\beta_5} \right) \right) \pm z_{1-\alpha/2} \text{ASE} \left[ f^{(1)}(\hat{\beta}) \right]
\]

where \( \text{ASE} \left[ f^{(1)}(\hat{\beta}) \right] = \sqrt{\text{Var} \left[ f^{(1)}(\hat{\beta}) \right]} \). Exponentiating the limits of (4.58) yields the analogous confidence interval for the odds ratio.

**4.10 An Alternative Approach to the Problem: The Independence Model**

The JMED model permits the formation of separate excess and deficiency curves, which naturally give way to an intersection point, notated as EPCP. In addition to the EPCP is also the probability of observing, on a given trial, a deficient condition or an excess condition. Ideally, it would be best if this probability were at a minimum.

If it is assumed that the basic outcome of an excess condition, E, on any given trial is independent of the basic outcome of a deficient condition, D, then this probability interest can be expressed as

\[
P(\text{SEV}) = P_{D \cup E}(Y_i = 1) = P_D(Y_i = 1) + P_E(Y_i = 1) - P_D(Y_i = 1)P_E(Y_i = 1)
\]

(4.59)
where

\[ P_{E}(Y_i = 1) = \frac{\exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}]}{1 + \exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}]} \]  

(4.60)

and

\[ P_{D}(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1x_{i1})}{1 + \exp(\beta_0 + \beta_1x_{i1})} \]  

(4.61)

We shall henceforth refer to the expression for \( P(\text{SEV}) = P_{\text{DOE}}(Y_i = 1) \) given in (4.59) as the Independence Model. The Independence Model is conditional on the logarithm base 10 concentration \( (x_{i1}) \) and the study identifier \( (x_{i2}) \).

Note that, as with the JMED model in Section 4.1, we again distinguish between the response, \( Y_i \) and the study identifier, \( x_{i2} \).

### 4.11 Investigating \( x_{\text{MIN DOE}} \)

The graphical representation of \( P(\text{SEV}) \) is a U-shaped curve as shown in Figure 2. The bottom of this curve represents the dose that will minimize the likelihood of severity. In the context of the copper data example, the bottom of the U is the exposure concentration that provides the lowest probability of experiencing adverse health effects resulting from deficient amounts of Cu, excess amounts of Cu, or both. This point shall be denoted as \( x_{\text{MIN DOE}} \).

Let

\[ P_{e_i} = P_{E}(Y_i = 1) = \frac{\exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}^*]}{1 + \exp[(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_{i1}^*]} \]  

(4.62)

\[ P_{d_i} = P_{D}(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1x_{i1}^*)}{1 + \exp(\beta_0 + \beta_1x_{i1}^*)} \]  

(4.63)
Figure 2: An example of a U-shape curve.

So

$$P(\text{SEV}) = P_{DJE}(Y_i = 1) = P_{e_i} + P_{d_i} - P_{e_i} \cdot P_{d_i}$$

To determine $x_{\text{MIN } DJE}$, we wish to find the value of $x_{t1}^{**}$ that minimizes $P(\text{SEV})$.

Set \[ \frac{\partial P(\text{SEV})}{\partial x_{t1}^{**}} = 0 \]

where

\[ \frac{\partial P(\text{SEV})}{\partial x_{t1}^{**}} = \frac{\partial}{\partial x_{t1}^{**}} [P_{e_i} + P_{d_i} - P_{e_i} \cdot P_{d_i}] = (\beta_1 + \beta_3)P_{e_i}(1-P_{e_i}) + \beta_1 P_{d_i}(1-P_{d_i}) \]

since

\[ \frac{\partial P_{e_i}}{\partial x_{t1}^{**}} = (\beta_1 + \beta_3)P_{e_i}(1 - P_{e_i}) \]
\[ \frac{\partial P_{di}}{\partial x_{i1}^{**}} = \beta_1 P_{di} (1 - P_{di}) \]

There is no closed form expression for the value of \( x_{i1}^{**} \) that minimizes \( P(\text{SEV}) \), so this problem will be solved numerically using the Newton Raphson Algorithm. The Newton Raphson Algorithm is an iterative procedure that updates the minimum at each step. In order to determine the value of \( x_{\text{MIN DUE}} \), a Newton step is computed as:

\[ x_{i1}^{**(n)} = x_{i1}^{**(n-1)} - \frac{f(x_{i1}^{**(n-1)})}{f'(x_{i1}^{**(n-1)})} \]

where

\[ f(x_{i1}^{**}) = (\beta_1 + \beta_3) P_{e_i} (1 - P_{e_i}) + \beta_1 P_{d_i} (1 - P_{d_i}) \] \hspace{1cm} (4.64)

and

\[ f'(x_{i1}^{**}) = (\beta_1 + \beta_3)^2 P_{e_i} (1 - P_{e_i}) + (\beta_1)^2 P_{d_i} (1 - P_{d_i}) \] \hspace{1cm} (4.65)

Since it is not possible to derive a closed form expression for \( x_{\text{MIN DUE}} \), a bootstrap approach to confidence interval estimation for \( x_{\text{MIN DUE}} \) must be employed. We describe this approach in Chapter 5. Note than an analogous bootstrap technique can also be applied to the confidence interval estimation of EPCP, as an alternative to the standard normal confidence interval. This is also discussed in Chapter 5.
4.12 The Odds of Excess or Deficiency at $x_{\text{MIN D},E}$

Of particular interest at the $x_{\text{MIN D},E}$ is the odds of excess or deficiency. Consider the logit function evaluated at $x_{\text{MIN D},E}$:

$$\text{logit} \left[ \pi (x_{\text{MIN D},E}) \right] = \log \left[ \frac{\pi (x_{\text{MIN D},E})}{1 - \pi (x_{\text{MIN D},E})} \right]$$

This represents an expression for the log odds of experiencing a severe health effect at $x_{\text{MIN D},E}$.

Consider two cases:

- $x_i^2 = 0$ : deficiency reading
- $x_i^2 = 1$ : excess reading

When $x_i^2 = 0$:

$$\log \left[ \frac{\pi (x_{\text{MIN D},E})}{1 - \pi (x_{\text{MIN D},E})} \right] = \beta_0 + \beta_1 x_{\text{MIN D},E} \quad (4.66)$$

$e^{\beta_0 + \beta_1 x_{\text{MIN D},E}}$ are the odds of experiencing a severe health effect caused by a deficient condition at $x_{\text{MIN D},E}$.

When $x_i^2 = 1$:

$$\log \left[ \frac{\pi (x_{\text{MIN D},E})}{1 - \pi (x_{\text{MIN D},E})} \right] = \beta_0 + \beta_2 + (\beta_1 + \beta_3) x_{\text{MIN D},E} \quad (4.67)$$

$e^{\beta_0 + \beta_2 + (\beta_1 + \beta_3) x_{\text{MIN D},E}}$ are the odds of experiencing a severe health effect caused by an excess condition at $x_{\text{MIN D},E}$. Note that since this log odds is conditional on $x_{\text{MIN D},E}$, we can also derive a standard normal confidence interval using the appropriate estimators and the same approach as for the EPCP.
4.13 The Independence Model Applied to the Species - Stratified JMED

The Independence Model will be applied to each of the models derived from the Species-stratified JMED model. For each species, a U-shape curve will be developed. The Newton Raphson algorithm will be used to find \( x_{\text{MIN DUE}} \) for each species. It is necessary to complete derivative calculations for each of the species.

4.13.1 Humans

For simplicity, denote:

\[
P_{e_i} = \frac{\exp((\beta_0 + \beta_2) + (\beta_1 + \beta_5)x_{i1})}{1 + \exp((\beta_0 + \beta_2) + (\beta_1 + \beta_5)x_{i1})} \quad (4.68)
\]

\[
P_{d_i} = \frac{\exp(\beta_0 + \beta_1x_{i1})}{1 + \exp(\beta_0 + \beta_1x_{i1})} \quad (4.69)
\]

So

\[
P(\text{SEV}) = P_{DUE}(Y_i = 1) = P_{e_i} + P_{d_i} - P_{e_i} \cdot P_{d_i} \quad (4.70)
\]

To minimize \( P(\text{SEV}) \):

Set \( \frac{\partial P(\text{Sev})}{\partial x_{i1}} = 0 \)

and solve for \( x_{i1} \) as a function of: \( \beta_0, \beta_1, \beta_2, \beta_5 \). That is,

\[
x_{i1} = f(\beta_0, \beta_1, \beta_2, \beta_5).
\]
\[ \frac{\partial P(\text{SEV})}{\partial x_{i1}} = \frac{\partial}{\partial x_{i1}} [P_{e_i} + P_{d_i} - P_{e_i} \cdot P_{d_i}] = (\beta_1 + \beta_5)P_{e_i}(1 - P_{e_i}) + \beta_1 P_{d_i}(1 - P_{d_i}) \]

since

\[ \frac{\partial P_{e_i}}{\partial x_{i1}} = (\beta_1 + \beta_5)P_{e_i}(1 - P_{e_i}) \]

and

\[ \frac{\partial P_{d_i}}{\partial x_{i1}} = \beta_1 P_{d_i}(1 - P_{d_i}) \]

Therefore,

\[ f(x_{i1}) = (\beta_1 + \beta_5)P_{e_i}(1 - P_{e_i}) + \beta_1 P_{d_i}(1 - P_{d_i}) \quad (4.71) \]

and

\[ f'(x_{i1}) = (\beta_1 + \beta_5)^2P_{e_i}(1 - P_{e_i}) + \beta_1^2 P_{d_i}(1 - P_{d_i}) \quad (4.72) \]

All necessary derivative calculations have been completed that are required to compute \( x_{\text{MIN DUJE}} \) for humans using the Newton Raphson algorithm.

### 4.13.2 Mice

Denote:

\[ P_{e_i} = \frac{\exp[(\beta_0 + \beta_2 + \beta_3 + \beta_6) + (\beta_1 + \beta_5 + \beta_6)x_{i1}]}{1 + \exp[(\beta_0 + \beta_2 + \beta_3 + \beta_6) + (\beta_1 + \beta_5 + \beta_6)x_{i1}]} \quad (4.73) \]
\[ P_{d_i} = \frac{\exp(\beta_0 + \beta_3 + (\beta_1 + \beta_6)x_{i1})}{1 + \exp(\beta_0 + \beta_3 + (\beta_1 + \beta_6)x_{i1})} \]  \hspace{1cm} (4.74)

So

\[ P(\text{SEV}) = P_{DUE}(Y_i = 1) = P_{e_i} + P_{d_i} - P_{e_i} \cdot P_{d_i} \]  \hspace{1cm} (4.75)

To minimize \( P(\text{SEV}) \):

Set \( \frac{\partial P(\text{SEV})}{\partial x_{i1}} = 0 \)

and solve for \( x_{i1} \) as a function of : \( \beta_0, \beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_8 \). That is,

\[ x_{i1} = f(\beta_0, \beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_8). \]

\[ \frac{\partial P(\text{SEV})}{\partial x_{i1}} = \frac{\partial}{\partial x_{i1}} [P_{e_i} + P_{d_i} - P_{e_i} \cdot P_{d_i}] = (\beta_1 + \beta_5 + \beta_6)P_{e_i}(1-P_{e_i}) + (\beta_1 + \beta_6)P_{d_i}(1-P_{d_i}) \]

since

\[ \frac{\partial P_{e_i}}{\partial x_{i1}} = (\beta_1 + \beta_5 + \beta_6)P_{e_i}(1 - P_{e_i}) \]

and

\[ \frac{\partial P_{d_i}}{\partial x_{i1}} = (\beta_1 + \beta_6)P_{d_i}(1 - P_{d_i}) \]

Therefore,

\[ f(x_{i1}) = (\beta_1 + \beta_5 + \beta_6)P_{e_i}(1 - P_{e_i}) + (\beta_1 + \beta_6)P_{d_i}(1 - P_{d_i}) \]  \hspace{1cm} (4.76)

and

74
\[ f'(x_{i1}) = (\beta_1 + \beta_5 + \beta_6)^2 P_e(1 - P_e) + (\beta_1 + \beta_6)^2 P_d(1 - P_d) \quad (4.77) \]

All necessary derivative calculations have been completed that are required to compute \( x_{\text{MIN D} \cup E} \) for mice using the Newton Raphson algorithm.

### 4.13.3 Rats

Denote:

\[ P_{e_i} = \frac{\exp[(\beta_0 + \beta_2 + \beta_4 + \beta_9) + (\beta_1 + \beta_5 + \beta_7) x_{i1}]}{1 + \exp[(\beta_0 + \beta_2 + \beta_4 + \beta_9) + (\beta_1 + \beta_5 + \beta_7) x_{i1}]} \quad (4.78) \]

\[ P_{d_i} = \frac{\exp(\beta_0 + \beta_4 + (\beta_1 + \beta_7) x_{i1})}{1 + \exp(\beta_0 + \beta_4 + (\beta_1 + \beta_7) x_{i1})} \quad (4.79) \]

So

\[ P(\text{SEV}) = P_{\text{DUE}}(Y_i = 1) = P_{e_i} + P_{d_i} - P_{e_i} \cdot P_{d_i} \quad (4.80) \]

To minimize \( P(\text{SEV}) \):

Set \( \frac{\partial P(\text{SEV})}{\partial x_{i1}} = 0 \)

and solve for \( x_{i1} \) as a function of : \( \beta_0, \beta_1, \beta_2, \beta_4, \beta_5, \beta_7, \beta_9 \). That is,

\[ x_{i1} = f(\beta_0, \beta_1, \beta_2, \beta_4, \beta_5, \beta_7, \beta_9). \]

\[ \frac{\partial P(\text{SEV})}{\partial x_{i1}} = \frac{\partial}{\partial x_{i1}} [P_{e_i} + P_{d_i} - P_{e_i} \cdot P_{d_i}] = (\beta_1 + \beta_5 + \beta_7) P_{e_i}(1 - P_{e_i}) + (\beta_1 + \beta_7) P_{d_i}(1 - P_{d_i}) \]
since

\[ \frac{\partial P_e}{\partial x_{i1}} = (\beta_1 + \beta_5 + \beta_r)P_e(1 - P_e) \]

and

\[ \frac{\partial P_d}{\partial x_{i1}} = (\beta_1 + \beta_r)P_d(1 - P_d) \]

Therefore,

\[ f(x_{i1}) = (\beta_1 + \beta_5 + \beta_r)P_e(1 - P_e) + (\beta_1 + \beta_r)P_d(1 - P_d) \quad (4.81) \]

and

\[ f'(x_{i1}) = (\beta_1 + \beta_5 + \beta_r)^2P_e(1 - P_e) + (\beta_1 + \beta_r)^2P_d(1 - P_d) \quad (4.82) \]

All necessary derivative calculations have been completed that are required to compute \( x_{\text{MIN-DUE}} \) for rats using the Newton Raphson algorithm.

### 4.14 The Odds at \( x_{\text{MIN-DUE}} \) for each Species

Consider the logit function evaluated at \( x_{\text{MIN-DUE}} \):

\[ \text{logit} \left[ \pi (x_{\text{MIN-DUE}}) \right] = \log \left[ \frac{\pi (x_{\text{MIN-DUE}})}{1 - \pi (x_{\text{MIN-DUE}})} \right] \]

This represents an expression for the log odds of a severe health condition.
at $x_{\text{MIN DUE}}$. Each species will have its own unique expression for the log odds evaluated at $x_{\text{MIN DUE}}$.

For each species, consider two cases:

- $x_{i2} = 0$ : deficiency reading
- $x_{i2} = 1$ : excess reading

### 4.14.1 Humans

When $x_{i2} = 0$:

$$
\log \frac{\pi (x_{\text{MIN DUE}})}{1 - \pi (x_{\text{MIN DUE}})} = \beta_0 + \beta_1 x_{\text{MIN DUE}} \quad (4.83)
$$

$e^{\beta_0 + \beta_1 x_{\text{MIN DUE}}}$ are the odds of experiencing a severe health effect attributed to a deficient reading at $x_{\text{MIN DUE}}$.

When $x_{i2} = 1$:

$$
\log \frac{\pi (x_{\text{MIN DUE}})}{1 - \pi (x_{\text{MIN DUE}})} = \beta_0 + \beta_2 + (\beta_1 + \beta_5) x_{\text{MIN DUE}} \quad (4.84)
$$

$e^{\beta_0 + \beta_2 + (\beta_1 + \beta_5) x_{\text{MIN DUE}}}$ are the odds of experiencing a severe health effect attributed to an excess reading at $x_{\text{MIN DUE}}$.

### 4.14.2 Mice

When $x_{i2} = 0$:

$$
\log \frac{\pi (x_{\text{MIN DUE}})}{1 - \pi (x_{\text{MIN DUE}})} = \beta_0 + \beta_3 + (\beta_1 + \beta_6) x_{\text{MIN DUE}} \quad (4.85)
$$

$e^{\beta_0 + \beta_3 + (\beta_1 + \beta_6) x_{\text{MIN DUE}}}$ are the odds of experiencing a severe health effect attributed to a deficient reading at $x_{\text{MIN DUE}}$. 
\( e^{\beta_0 + \beta_3 + (\beta_1 + \beta_6)x_{\text{MIN } D\cup E}} \) are the odds of experiencing a severe health condition attributed to a deficient reading at \( x_{\text{MIN } D\cup E} \).

When \( x_{i2} = 1 \):

\[
\log \left[ \frac{\pi (x_{\text{MIN } D\cup E})}{1 - \pi (x_{\text{MIN } D\cup E})} \right] = \beta_0 + \beta_2 + \beta_3 + \beta_8 + (\beta_1 + \beta_5 + \beta_6)x_{\text{MIN } D\cup E} \tag{4.86}
\]

\( e^{\beta_0 + \beta_3 + \beta_8 + (\beta_1 + \beta_5 + \beta_6)x_{\text{MIN } D\cup E}} \) are the odds of experiencing a severe health condition attributed to an excess reading at \( x_{\text{MIN } D\cup E} \).

4.14.3 Rats

When \( x_{i2} = 0 \):

\[
\log \left[ \frac{\pi (x_{\text{MIN } D\cup E})}{1 - \pi (x_{\text{MIN } D\cup E})} \right] = \beta_0 + \beta_4 + (\beta_1 + \beta_7)x_{\text{MIN } D\cup E} \tag{4.87}
\]

\( e^{\beta_0 + \beta_4 + (\beta_1 + \beta_7)x_{\text{MIN } D\cup E}} \) are the odds of experiencing a severe health condition attributed to a deficient reading at \( x_{\text{MIN } D\cup E} \).

When \( x_{i2} = 1 \):

\[
\log \left[ \frac{\pi (x_{\text{MIN } D\cup E})}{1 - \pi (x_{\text{MIN } D\cup E})} \right] = \beta_0 + \beta_2 + \beta_4 + \beta_9 + (\beta_1 + \beta_5 + \beta_7)x_{\text{MIN } D\cup E} \tag{4.88}
\]

\( e^{\beta_0 + \beta_2 + \beta_4 + \beta_9 + (\beta_1 + \beta_5 + \beta_7)x_{\text{MIN } D\cup E}} \) are the odds of experiencing a severe health condition attributed to an excess reading at \( x_{\text{MIN } D\cup E} \).
5 Simulation Study

To investigate the validity of the methodology described in Chapter 4, we present a simulation study. For simplicity, we focus on the JMED model without species stratification.

5.1 Validation Procedure for EPCP

Specifically, we consider working with the model given in equation (4.1); namely,

\[
\pi_i = P(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2})}
\]

with \(x_{i1}\) representing the log_{10} concentration of the \(i^{th}\) observation and \(x_{i2}\) an indicator variable for excess or deficiency, and where

\[
\beta_0 = -2.5164
\]
\[
\beta_1 = -3.4780
\]
\[
\beta_2 = 1.4909
\]
\[
\beta_3 = 4.5053
\]

yielding \(EPCP = \frac{\beta_2}{\beta_3} = -0.3309\). The values comprising \(\beta\) coincide with the estimate of the \(\beta\) parameter vector obtained when fitting the above model to the data set that will be considered in Chapter 6. This data set consists of 3886 observations; 1943 are from experiments on deficiency and 1943
are from experiments on excess. Each observation also had available a log concentration measure and a species identifier.

Initially, we generated a single simulated dataset by computing \( \pi_i \), given in the model above, for each observation in the actual data set that is available for analysis in Chapter 6. For a given observation \( i \), a simulated response, \( Y^{(s)}_i \), is generated from a Bernoulli distribution with parameter \( \pi_i \). The generated \( Y^{(s)}_i \) along with the values for \( x_{i1} \) and \( x_{i2} \) in the data set described above were used to fit to the model in equation (4.1), providing \( \hat{\beta}^{(s)}_0, \hat{\beta}^{(s)}_1, \hat{\beta}^{(s)}_2, \hat{\beta}^{(s)}_3 \), and an associated estimated asymptotic covariance matrix. These estimates then allowed for the computation of a point estimate for EPCP, \( \text{EPCP}^{(s)} \), along with an approximate 95% large sample confidence interval using the formula in equation (4.10).

This procedure was repeated 1000 times, providing 1000 point estimates of the EPCP and their associated normal confidence intervals. Over the 1000 replications, the average \( \text{EPCP}^{(s)} \) was -0.3299609 which compares very favourably to the true EPCP of -0.3309. In addition, we determined the proportion of 95% confidence intervals that covered the true EPCP to be 0.954, which is extremely close to the nominal rate.

Note that an alternative approach for interval estimation of EPCP could be through the use of a bootstrap methodology. For example, if we consider the \( s^{th} \) simulated data set described above with its associated model fit that provided \( \hat{\beta}^{(s)}_0, \hat{\beta}^{(s)}_1, \hat{\beta}^{(s)}_2, \hat{\beta}^{(s)}_3 \), we would proceed with computing \( \pi_i^{(s)} \) for all \( i \). For each given observation \( i \), a bootstrap response, \( Y^{(s)(b)}_i \), is generated from a Bernoulli distribution with parameter \( \pi_i^{(s)} \). The generated \( Y^{(s)(b)}_i \) along with the values for \( x_{i1} \) and \( x_{i2} \) in the data set described above are then used
to fit the model in equation (4.1), providing \( \hat{\beta}_0^{(s)} \), \( \hat{\beta}_1^{(s)} \), \( \hat{\beta}_2^{(s)} \), and \( \hat{\beta}_3^{(s)} \). These estimates are employed to compute the estimate \( \hat{\text{EPCP}}^{(s)}_{(b)} \). For the \( s^{th} \) simulated data set, if we repeat this procedure for \( b = 1, \ldots, B \) bootstrap samples, we can determine \( B \) estimates \( \hat{\text{EPCP}}^{(s)}_{(b)} \), which can be used to determine a 95% bootstrap confidence interval for EPCP given by

\[
\left( \hat{\text{EPCP}}^{(s)}_{[0.025]}, \hat{\text{EPCP}}^{(s)}_{[0.975]} \right)
\]

where \( \hat{\text{EPCP}}^{(s)}_{[0.025]} \) and \( \hat{\text{EPCP}}^{(s)}_{[0.975]} \) represent the 2.5th and 97.5th quantiles of the empirical distribution of the \( B \) values for \( \hat{\text{EPCP}}^{(s)}_{(b)} \).

For each of the \( s = 1, \ldots, 1000 \) data sets that we simulated, we generated \( b = 1, \ldots, 1000 \) bootstrap samples, using the latter to compute a 95% bootstrap confidence interval for EPCP for each simulated data set. We determined the proportion of 95% bootstrap confidence intervals that covered the true EPCP to be 0.945, which is very close to the nominal rate. Note also that the standard normal confidence intervals and the bootstrap confidence intervals based on the empirical distribution were very close (95.4% and 94.5%), justifying the use of the normal approximation.

5.2 Investigating the Limitations of the EPCP

The bias and variance approximations given in (4.7) and (4.9) for the \( \hat{\text{EPCP}} \) are greatly impacted by \( \beta_3 \) and \( \text{Var} [\beta_3] \). Limitations of the EPCP will depend on these values. When the bias and variance of the EPCP are inflated, the reliability of any inference performed on the EPCP is compromised. In order to investigate the impact of the size of \( \beta_3 \) relative to the other \( \beta \) parameters in
the model, we conducted another simulation study. Specifically, we repeated
the simulation described in Section 5.1 where the parameters $\beta_0, \beta_1, \beta_2$ were
set to -2, -0.5, and 1, respectively, and we used different values of $\beta_3$ ranging
from 0.6 to 3. These values were selected so that $\beta_1 + \beta_3 > 0$. The results are
summarized in Table 2. Note that the coverage rates provided are for 95%
confidence intervals based on a bootstrap approach.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
$\beta_3$ & EPCP & Bias[EPCP] & Var[EPCP] & Coverage Rate(%) \\
\hline
0.6 & -1.66 & -0.027 & 0.0434 & 95.3 \\
1 & -1 & -0.00314 & 0.0094 & 94.8 \\
1.5 & -0.66 & -0.00055 & 0.0056 & 96.0 \\
2 & -0.5 & -0.00049 & 0.0045 & 94.9 \\
2.5 & -0.4 & -0.00044 & 0.0033 & 93.6 \\
3 & -0.333 & -0.00041 & 0.0025 & 95.2 \\
\hline
\end{tabular}
\caption{Performance of EPCP for various values of $\beta_3$.}
\end{table}

Since 1000 replications were run for each chosen value of $\beta_3$, the Monte-
Carlo simulation error on the 95% bootstrap intervals can be approximated
as $3\sqrt{\frac{0.95(0.05)}{1000}} = 0.02$. This indicates that the proportion of intervals that
contain the true EPCP should lie between (0.93, 0.97). We do not detect
any evidence of a lack of proper coverage for any of these $\beta_3$ values.

\section{Validation Procedure for $x_{\text{MIN DUE}}$}

In order to validate the estimation procedure for $x_{\text{MIN DUE}}$ presented in Chap-
ter 4, we work with the Independence Model given in (4.59).

\begin{equation*}
P(\text{SEV}) = P_{DUE}(Y_i = 1) = P_D(Y_i = 1) + P_E(Y_i = 1) - P_D(Y_i = 1) \cdot P_E(Y_i = 1)
\end{equation*}
where
\[
P_E(Y_i = 1) = \frac{\exp(-1.0215 + 1.0274x_{i1})}{1 + \exp(-1.0215 + 1.0274x_{i1})}
\]
\[
P_D(Y_i = 1) = \frac{\exp(-2.5164 - 3.478x_{i1})}{1 + \exp(-2.5164 - 3.478x_{i1})}
\]

The values of \(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2,\) and \(\hat{\beta}_3\) agree with those presented in the validation procedure for EPCP. Also, the same data set of size 3886 was used, comprised of 1943 excess experiments and 1943 deficiency experiments.

In order to find \(x_{\text{MIN D\cup E}}\), we must minimize equation (4.59) with respect to \(x_{i1}\). We use the Newton Raphson method since there is no closed form solution for \(x_{i1}\). The algorithm was initialized with \(x_{i1} = 0.1\) and the convergence criteria was selected as \(\epsilon = 1 \times 10^{-08}\). The algorithm converged when \(x_{i1} = -0.00971661\). Therefore, \(x_{\text{MIN D\cup E}} = -0.00971661\).

In order to assess the procedures proposed in Chapter 4 for estimating \(x_{\text{MIN D\cup E}}\), as in the case of the EPCP, we generated a single simulated data set using equation (4.1) by computing \(\pi_i\) for each observation available for analysis in Chapter 6. For a given observation, \(i\), a simulated response, \(Y_i^{(s)}\), is generated from a Bernoulli distribution with parameter \(\pi_i\). The generated \(Y_i^{(s)}\) along with the values for \(x_{i1}\) and \(x_{i2}\) were fit to the model in equation (4.1), providing \(\hat{\beta}_0^{(s)}, \hat{\beta}_1^{(s)}, \hat{\beta}_2^{(s)},\) and \(\hat{\beta}_3^{(s)}\). This procedure was repeated 1000 times, to generate 1000 simulated \(\beta\) parameters. For each of these \(s = 1, \ldots, 1000\) simulated data sets, we can get an estimate \(\hat{x}_{x_{\text{MIN D\cup E}}}^{(s)}\). The average of these 1000 estimates for \(\hat{x}_{x_{\text{MIN D\cup E}}}^{(s)}\) is -0.0103, which is close to the true value of \(x_{\text{MIN D\cup E}}\).

Unlike EPCP, it is not possible to write a closed form expression for \(x_{\text{MIN D\cup E}}\). Thus, in order to determine a confidence interval for the latter
quantity, a bootstrap technique analogous to the one used for EPCP is necessary.

Suppose we consider one of the 1000 simulated data sets described above with its associated model fit that provides $\hat{\beta}_0^{(s)}, \hat{\beta}_1^{(s)}, \hat{\beta}_2^{(s)}, \text{and} \hat{\beta}_3^{(s)}$. We may proceed by computing $\pi_i^{(s)}$. For each observation, $i$, a bootstrap response, $Y_i^{(s)(b)}$, is generated from a Bernoulli distribution with parameter $\pi_i^{(s)}$. The generated $Y_i^{(s)(b)}$, along with the values for $x_{i1}$ and $x_{i2}$ were fit to the model in equation (39), providing $\hat{\beta}_0^{(s)(b)}, \hat{\beta}_1^{(s)(b)}, \hat{\beta}_2^{(s)(b)}, \text{and} \hat{\beta}_3^{(s)(b)}$. To obtain one bootstrap estimate of $x_{\text{MIN DUE}}$, denoted as $\hat{x}_{\text{MIN DUE}}^{(s)(b)}$, we applied the Newton Raphson algorithm to minimize $P(\text{SEV})$ evaluated at $\hat{\beta}_0^{(s)(b)}, \hat{\beta}_1^{(s)(b)}, \hat{\beta}_2^{(s)(b)}, \text{and} \hat{\beta}_3^{(s)(b)}$.

Within one simulated data set, we generated 1000 bootstrap samples, and determine $\hat{x}_{\text{MIN DUE}}^{(s)(b)}$, $b = 1, \ldots, 1000$ for each. Using these estimates, we can obtain a 95% bootstrap confidence interval for $x_{\text{MIN DUE}}$; namely

$$\left(\hat{x}_{\text{MIN DUE}}^{(s)(b)}; \hat{x}_{\text{MIN DUE}}^{(s)(b)}, [0.025], \hat{x}_{\text{MIN DUE}}^{(s)(b)}, [0.975]\right)$$

where $\hat{x}_{\text{MIN DUE}}^{(s)(b)}$ and $\hat{x}_{\text{MIN DUE}}^{(s)(b)}$ represent the 2.5th and 97.5th quantiles of the empirical distribution of the 1000 values for $\hat{x}_{\text{MIN DUE}}^{(s)(b)}$. For each of the $s = 1, \ldots, 1000$ data sets that we simulated, we generated $b = 1, \ldots, 1000$ bootstrap samples, using the latter to compute a 95% bootstrap confidence interval for $x_{\text{MIN DUE}}$ for each simulated data set.

We determined the proportion of these 95% bootstrap confidence intervals that cover $x_{\text{MIN DUE}}$ to be 0.942, which is very close to the nominal rate.
5.4 Summary

To summarize, the simulation study presented here has demonstrated the EPCP and $x_{\text{MIN DUE}}$ estimators perform well for sufficiently large samples. Since the coverage of the standard normal and bootstrap confidence intervals are very close to the nominal rate, we can be assured that the EPCP and $x_{\text{MIN DUE}}$ are reliable estimators. The reliability of these estimators should also extend to the species-stratified case. The limitations of the JMED model have also been discussed, but we have shown that the inference techniques proposed here for EPCP and $x_{\text{MIN DUE}}$ should be reliable for the copper database.
6 Application with Copper Database

In this chapter, we discuss the modifications made to the Copper(Cu) Database that prepared it for the application of logistic regression, also presented here. Many of these modifications were implemented by Chambers et al. (2010). The results from the JMED and Independence models will be interpreted within the framework of health risk and toxicity.

In preparation for the application of categorical regression, the database received minor modifications. Since data from multiple species were combined, Chambers et al. (2010) redefined the dose metric to accommodate interspecies differences. To create a universal measure, exposure concentration was expressed in terms of body weight, namely, mg/kg bw/day. In the original database developed in 2002, severity scores were assigned to incidence or group level data, where groups were classified according to size (level 1, 2, or 3). Within each group, severity scores were assumed to be cumulative; those effects listed in lower severity categories are presumed to also occur in the higher categories. The highest severity score observed was assigned to the entire group. Instead of using weight factors to reflect various group sizes in the regression analysis, the articles in the database were re-examined and the exact group size from each experiment was recorded. The group level observations have been replicated appropriately so that each record in the database represents one subject. Converting group data into individualized data expanded the database and provided more information available for analysis.

Exploratory analysis with an ordinal response variable presented chal-

86
The Cu database contains insufficient amounts of information to build models for the excess and deficiency data sets using an ordinal response variable. For consistency, the response variable was dichotomized to represent a Cu reading within the homeostatic range, or a Cu reading outside the homeostatic range. Originally, the $i^{th}$ observation in the Cu database was assigned a severity score, $S_i$, where $S_i \in \{-6, -5, \ldots, 5, 6\}$. Observations with severity scores that satisfy $|S_i| \leq 1$ are considered to be within the homeostatic range and are recoded as $Y_i = 0$. On the contrary, severity scores that satisfy $|S_i| \geq 2$ are considered to lie outside the homeostatic range and are recoded as $Y_i = 1$. Both data sets contain enough information to support the construction of a model with a binary response variable.

The Cu database contains a large collection of independent variables available for analysis. To establish which explanatory variables were significant in predicting severity outcome, a stepwise regression exercise was conducted. To begin, several binary logistic regression models were fit, each incorporating one independent variable as the main effect. If the main effect was significant at $\alpha = 0.05$, a second independent variable was added. If the main effect was insignificant, the stepwise regression procedure did not continue. All possible models were constructed, with all combinations of independent variables accounted for. Exposure concentration was the only continuous main effect that was significant in predicting severity. To extend this model, responses were also stratified by species.

In an effort to ensure excess and deficiency observations equally contributed to the model building process, 1943 observations from the excess data set were randomly selected and combined with the 1943 deficiency ob-
servations to create a data set of size 3886. We illustrate the methodology proposed in Chapter 4 on this data set. Specifically, we consider models with and without species stratification. The latter, presented in Sections 6.1 through 6.4, and 6.6, is simply included as an illustration of the suggested approaches, since species confounding will compromise the ability to draw any health-related conclusions for humans.

6.1 Representations of the JMED

Fitting the JMED model in (4.1) to the copper data set of 3886 observations described earlier yields $\hat{\beta}_0 = -2.516, \hat{\beta}_1 = -3.478, \hat{\beta}_2 = 1.4909$ and $\hat{\beta}_3 = 4.505$, along with an associated estimated covariance matrix. A plot of the excess and deficiency curves resulting from this fit is given in Figure 3.

![Excess and Deficiency Logistic Regression Curves](image)

Figure 3: The components of the JMED: Excess and Deficiency curves.
The deficiency curve exhibits a decreasing relationship, where the odds of a deficient condition decrease with increasing concentration level. Exposure concentrations considered to be Cu deficient provide a high probability of experiencing adverse health effects. As exposure concentration increases, the chances of severity decrease. In contrast, the excess curve exhibits a reverse relationship. When the exposure concentration is small, the probability of experiencing adverse health effects is very small. As exposure concentration increases, the individual is more likely to experience adverse health effects.

The slope parameters of the excess and deficiency curves demonstrate the differential behaviour exhibited by these two data sets. The excess model has a relatively small slope parameter estimate ($\hat{\beta}_1 = 1.0274$), while the deficiency model has a relatively large slope parameter estimate ($\hat{\beta}_1 = -3.478$). For an excess condition, as the dose of Cu increases, the probability of severity gradually increases. For a deficiency condition, as the dose of Cu increases, the probability of severity drastically decreases.

6.2 Interpreting the EPCP

The EPCP is estimated by $-\frac{\hat{\beta}_2}{\hat{\beta}_3} = -\frac{1.4909}{3.5053} = -0.3309$. When an individual ingests $10^{-0.3309} = 0.46673$ mg/kg bw/day, the probability of experiencing adverse health effects from excess exposure is equivalent to the probability of experiencing adverse health effects from deficient exposure. A daily exposure concentration of 0.46673 mg/kg bw will protect an individual from a harmful ‘unknown’ Cu condition. In addition, using equation (4.10), we can compute a 95% confidence interval for $10^{\text{EPCP}}$. Appropriately transforming the upper
and lower limits of this interval allows us to conclude with 95% confidence, this exposure concentration lies between 0.41839 and 0.53042 mg/kg bw/day.

6.3 Interpreting the Odds of Severity at the EPCP

Using equation (4.12), the log odds of severity at the EPCP are estimated as \( \hat{\beta}_0 - \frac{\hat{\beta}_1 \hat{\beta}_2}{\hat{\beta}_3} = -2.516 + \frac{3.478(1.4909)}{4.505} = -1.3634 \), which suggests that the odds of severity are estimated by \( e^{-1.3634} = 0.2557 \). At the EPCP, we estimate a severe health effect will occur \( \frac{25.57}{100-25.57} = 34.59 \% \) of the time, while a non-severe health effect will occur \( 1 - \frac{25.57}{100-25.57} = 65.41 \% \) of the time. At the EPCP, the odds are in favour of not experiencing a severe health effect. In addition, using equation (4.18), we can compute a 95% confidence interval for the log odds at EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, the odds lie between 0.2173 and 0.2998.

6.4 Odds Ratio of Excess to Deficiency at a Specified \( \log_{10} \) (Concentration)

Table 3 summarizes the estimates of the odds ratios of excess to deficiency, given in equation (4.22) for log concentrations in the range \((-3, 2)\) in increments of 0.5. In addition, using equation (4.23), we estimate a standard normal 95% confidence interval at each log concentration value. The precision of the estimate of the odds ratio is determined by the width of its confidence interval. A wide confidence interval indicates a less precise estimate.

For example, when \( \log_{10} \) concentration is equal to -3, the dose is \( 10^{-3} = \)
0.001 mg/kg bw/day. When the exposure concentration is 0.001 mg/kg bw, the odds of an excess condition are $5.99e^{-06}$ times the odds of a deficiency condition. In other words, a deficiency condition is much more likely than an excess condition. This observation is true for all log$_{10}$ concentration values less than 0, or equivalently all dose levels less than 0.32 mg/kg bw. The odds ratio estimates do increase however, as the log$_{10}$ concentration approaches zero. At the same time, the 95% confidence intervals are more narrow.

Suppose we consider the case when log$_{10}$ concentration is equal to 0 and the dose is $10^0 = 1$ mg/kg bw. When the exposure concentration is 1 mg/kg bw, the odds of an excess condition are 4.44 times the odds of a deficient condition. For all log$_{10}$ concentration values greater than or equal to 0 with corresponding exposure concentrations greater than 1 mg/kg bw, the odds favour an excess condition. These odds ratios increase significantly as the concentration level increases, along with the width of the analogous 95% confidence intervals.
6.5 Stratification by Species

After the database was expanded to convert group level data to individual level data, studies conducted on mice and rats comprised an overwhelming proportion of the database. In order to perform a regression analysis focused on stratification by species, data sets well-suited for logistic regression were created by removing any observations considered to be ‘noise.’ It was necessary to delete replicates, so as to ensure the independence of observations in the data set to be analyzed. Interpreting results in the context of toxicity and health is the primary focus.

Fitting the JMED model in (4.24) to the individual level human, rat and mice data described above yields $\hat{\beta}_0 = -9.513036$, $\hat{\beta}_1 = -4.654968$, $\hat{\beta}_2 = 15.9496$, $\hat{\beta}_3 = 6.436742$, $\hat{\beta}_4 = 3.416551$, $\hat{\beta}_5 = 11.58433$, $\hat{\beta}_6 = -4.893107$, $\hat{\beta}_7 = -4.833012$, $\hat{\beta}_8 = -17.46571$, and $\hat{\beta}_9 = -13.1585$ along with an associated variance-covariance matrix. A plot of excess and deficiency curves is provided for each species in Figure 4.

![Excess and Deficiency Curves](image)

Figure 4: Species stratified models for Cu Database.
This figure demonstrates humans are most sensitive to excess and deficient Cu conditions, followed by rats, then mice. For rats, mice and humans, the slope coefficient for deficiency is larger than excess. A large slope coefficient for deficiency indicates that as the exposure concentration increases, the probability of severity decreases rapidly. The smaller excess slope coefficient means that as concentration increases, the probability of severity increases gradually.

### 6.5.1 Interpreting EPCP\(_{HU}\), EPCP\(_{MU}\), and EPCP\(_{RT}\)

#### Humans

The EPCP is estimated by \(\frac{\dot{\beta}_2}{\dot{\beta}_5} = -\frac{15.95}{11.58} = -1.377\). For humans, at an exposure concentration of \(10^{-1.377} = 0.042\) mg/kg bw, the probability of experiencing adverse health effects from excess exposure is equivalent to the probability of experiencing adverse health effects from deficient exposure. In addition, using equation (4.10), we can compute a 95% confidence interval for EPCP\(_{HU}\). Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, \(10^{EPCP_{HU}}\) lies between 0.0033 and 0.4619 mg/kg bw.

#### Mice

The EPCP is estimated by \(\frac{\dot{\beta}_2 + \dot{\beta}_8}{\dot{\beta}_5} = -\frac{15.95 - 17.466}{11.58} = 0.1308\). For mice, at an exposure concentration of \(10^{0.1308} = 1.3516\) mg/kg bw, the probability of experiencing adverse health effects from excess exposure is equivalent to the probability of experiencing adverse health effects from deficient exposure.
addition, using equation (4.30), we can compute a 95% confidence interval for \( \text{EPCP}_{\text{MU}} \). Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, \( 10^{\text{EPCP}_{\text{MU}}} \) lies between 0.1288 and 14.12 mg/kg bw.

**Rats**

The EPCP is estimated by \( \frac{\hat{\beta}_2 + \hat{\beta}_9}{\hat{\beta}_5} = \frac{-15.95 - 13.16}{11.58} = -0.2409 \). For rats, at an exposure concentration of \( 10^{-0.2409} = 0.5742 \) mg/kg bw, the probability of experiencing adverse health effects from excess exposure is equivalent to the probability of experiencing adverse health effects from deficient exposure. In addition, using equation (4.30), we can compute a 95% confidence interval for \( \text{EPCP}_{\text{RT}} \). Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, \( 10^{\text{EPCP}_{\text{RT}}} \) lies between 0.0433 and 7.709 mg/kg bw.

For the non-stratified model, the estimate \( 10^{\hat{\text{EPCP}}} \) was found to be 0.46673 mg/kg bw. Analogous estimates from the stratified model were 0.03905, 1.3516, and 0.5742 mg/kg bw for humans, mice, and rats, respectively. Note that the EPCP for the non-stratified model is fairly close to \( \text{EPCP}_{\text{RT}} \). The non-stratified model is largely driven by the rat data set. Integrating species into the model has highlighted the differences present in the exposure-response behaviour displayed by humans, mice, and rats. The probability of severity on the vertical axis associated with the EPCP differs across species. Figure 4 showcases the similar behaviour exhibited by rats and mice, and the differential tendencies exhibited by humans.
6.5.2 Interpreting the Log Odds of Severity at EPCP$_{HU}$, EPCP$_{MU}$, and EPCP$_{RT}$

The log odds evaluated at the EPCP provides a measure to assess the odds of a severe condition where it is least likely.

**Humans**

Using equation (4.12), the log odds of severity at the EPCP are estimated as

\[
\hat{\beta}_0 - \frac{\hat{\beta}_1 \hat{\beta}_2}{\hat{\beta}_3} = -9.513 - \frac{(-4.65)(15.95)}{11.38} = -3.104,
\]

which suggests the odds are estimated by \(e^{-3.104} = 0.04487\). For a human, there is a \(\frac{4.7}{95.3}\%\) chance of experiencing a severe health effect when a dose of 0.03905 mg/kg bw is administered, while there is a \(1 - \frac{4.7}{95.3}\%\) chance of no severe health effect. At the EPCP, while the odds are in favour of not experiencing a severe health effect, the probability of severity at this exposure level appears somewhat high. However, it is unclear as to how the data used to obtain this estimate was collected, and whether it is representative of the general population as a whole.

Using equation (4.18), with \(C^T\) defined in (4.33), we can compute a 95% confidence interval for the log odds at the EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, that the odds lie between \(5.75e^{-8}\) and \(3.50e^4\). This wide confidence interval reflects the variability in the data.
Mice

Using equation (4.34), the log odds of severity at the EPCP are estimated as
\[ \hat{\beta}_0 + \hat{\beta}_3 - \frac{(\hat{\beta}_1 + \hat{\beta}_4)(\hat{\beta}_2 + \hat{\beta}_5)}{\hat{\beta}_5} = -9.513 + 6.437 - \frac{(-4.65 - 4.893)(15.95 - 17.466)}{11.58} = -4.326, \]
which suggests the odds are 0.01322. For a mouse, there is a \( \frac{1.322}{98.678} = 1.34\% \) chance of experiencing a severe health effect. At the EPCP, the odds are in favour of not experiencing a severe health effect, while there is a \( 1 - \frac{1.322}{98.678} = 98.866\% \) chance of no severe health effect. In addition, using equation (4.41), we can compute a 95% confidence interval for the log odds at the EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, that the odds lie between 0.00053 and 0.3276.

Rats

Using equation (4.35), the log odds of severity at the EPCP are estimated as
\[ \hat{\beta}_0 + \hat{\beta}_4 - \frac{(\hat{\beta}_1 + \hat{\beta}_7)(\hat{\beta}_2 + \hat{\beta}_8)}{\hat{\beta}_8} = -9.513 + 3.416 - \frac{(-4.65 - 4.833)(15.95 - 13.158)}{11.58} = -3.81, \]
which suggests the odds are 0.022. For a rat, there is a \( \frac{2.2}{97.8} = 2.25\% \) chance of experiencing a severe health effect when a dose of 0.5742 mg/kg bw is administered, while there is a \( 1 - \frac{2.2}{97.8} = 97.85\% \) chance of no severe health effect. In addition, using equation (4.41), we can compute a 95% confidence interval for the log odds at the EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, that the odds lie between 0.00052 and 0.944.
6.5.3 Interpreting Species-to-Species Odds Ratios at the EPCP

Species-to-species odds ratios compare the odds of severity between two species at their respective EPCP’s. Using these odds ratios, we can establish which species is more sensitive to a severe health effect at its EPCP.

Mice to Humans

Using equation (4.45), the odds ratio for mice to humans is estimated as $e^{\hat{\beta}_3 - \hat{\beta}_2 \hat{\beta}_4 + \hat{\beta}_1 \hat{\beta}_6 + \hat{\beta}_5 \hat{\beta}_8} = 0.29465$. At doses of 1.3516 mg/kg bw and 0.03905 mg/kg bw for mice and humans, respectively, mice are 0.3 times more likely to exhibit a non-normal condition. Therefore, humans are more sensitive to a non-normal condition than mice. In addition, using equation (4.51), we can compute a 95% confidence interval for the odds ratio. With 95% confidence, we estimate that the odds ratio lies between -7.806 and 8.396.

Rats to Humans

Using equation (4.46), the odds ratio for rats to humans is estimated as $e^{\hat{\beta}_4 - \hat{\beta}_2 \hat{\beta}_7 + \hat{\beta}_1 \hat{\beta}_9 + \hat{\beta}_5 \hat{\beta}_9} = 0.4944$. At doses of 0.5742 mg/kg bw and 0.03905 mg/kg bw for rats and humans, respectively, rats are 0.49 times more likely to exhibit a non-normal condition. Therefore, humans are more sensitive to a non-normal condition than rats. In addition, using equation (4.51), we can compute a 95% confidence interval for the odds ratio. With 95% confidence, we estimate that the odds ratio lies between -4.735 and 5.721.
Rats to Mice

Using equation (4.52), the odds ratio for rats to mice is estimated as 

\[ e^{\hat{\beta}_4 - \hat{\beta}_3 - (\hat{\beta}_2 + \hat{\beta}_9)(\hat{\beta}_1 + \hat{\beta}_7) + (\hat{\beta}_2 + \hat{\beta}_8)(\hat{\beta}_1 + \hat{\beta}_6)} = 0.1375. \]

At doses of 0.5742 mg/kg bw and 1.3516 mg/kg bw for rats and mice, respectively, rats are 0.14 times more sensitive to a non-normal condition than mice. Therefore, mice are more sensitive to a non-normal condition than rats. In addition, using equation (4.58), we can compute a 95% confidence interval for the odds ratio. With 95% confidence, we estimate that the odds ratio lies between -4.012 and 4.287.

6.6 An Alternative Approach: \( x_{\text{MIN DUE}} \)

The JMED model fit in Section 6.1 allows for the estimation of \( P(\text{SEV}) \) in (4.59) at different levels of log concentration. A plot is presented in Figure 5, along with the excess and deficiency curves in Figure 3.

Using the Newton Raphson approach described in Section 4.11, \( x_{\text{MIN DUE}} \) is estimated by -0.009719. For all species, a dose of \( 10^{-0.009719} = 0.97787 \) mg/kg bw will minimize the likelihood of experiencing adverse health effects emerging from a deficient condition, an excess condition, or both. In addition, using the methodology described in Section 5.3, we can initially construct a 95% bootstrap confidence interval for \( x_{\text{MIN DUE}} \). Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, \( 10^{x_{\text{MIN DUE}}} \) lies between 0.8166 and 1.044 mg/kg bw.
Figure 5: A display of the U-shape curve generated by modeling P(Sev).

6.6.1 Interpreting the Odds of Excess or Deficiency at $x_{\text{MIN DUE}}$

Using equation (4.67), the log odds of an excess condition at $x_{\text{MIN DUE}}$ are estimated by

$$\hat{\beta}_0 + \hat{\beta}_2 + (\hat{\beta}_1 + \hat{\beta}_3)\hat{x}_{\text{MIN DUE}} = -2.516 + 1.4909 + (-3.478 + 4.505)(-0.009719) = -1.0351,$$

which suggests the odds are $e^{-1.035} = 0.355$. There is a $\frac{35.5}{94.5} = 35.04\%$ chance that a dose of 0.98 mg/kg bw/day will lead to severe health effects.

Using equation (4.66), the log odds of a deficiency condition at $x_{\text{MIN DUE}}$ are estimated by

$$\hat{\beta}_0 + \hat{\beta}_1\hat{x}_{\text{MIN DUE}} = -2.516 + (-3.478(-0.009719)) = -2.482,$$

which suggests the odds are $e^{-2.482} = 0.08355$. Assuming a deficiency condition, there is a $\frac{8.355}{91.65} = 9.11\%$ chance that a dose of 0.98 mg/kg bw/day will lead to severe health effects. When a dose of 0.98 mg/kg bw/day is administered, the odds suggest a non-normal reading caused by an excess condition is more likely.
6.7 The Independence Model Applied to the Species-Stratified JMED

As an extension of the Species-Stratified JMED, a model is created for each species that represents the probability of severity from an excess reading, a deficient reading, or both, using the Independence Model. Using the fit of the JMED in (4.24), for each species, we computed a point estimate and bootstrap confidence interval for the species-specific $x_{\text{MIN D} \cup E}$, along with estimates of the odds of deficient and excess conditions at this log concentration. We also present plots of $P(\text{SEV})$ for each species, given in (4.70), (4.75), and (4.80).

6.7.1 Humans

Estimating $x_{\text{MIN D} \cup E}$

A plot of $P(\text{SEV})$ versus log concentration is given in Figure 6. For humans, $x_{\text{MIN D} \cup E}$ is estimated by -1.408 using the Newton Raphson approach described in Section 4.11. A dose of $10^{-1.408} = 0.03905 \text{ mg/kg bw}$ will minimize the probability of severity emerging from an excess or deficient condition, or both. In addition, using the methodology described in Section 5.3, we can construct a 95% bootstrap confidence interval for $x_{\text{MIN D} \cup E}$. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, $10^{x_{\text{MIN D} \cup E}}$ lies between 0.02253 and 0.06375 mg/kg bw. This confidence interval is displayed in Figure 6, along with the exposure concentration and severity response for each of the observations used in the
model fit. The interval can be viewed as a range of acceptable values for dose that minimize the probability of severity. A similar conclusion can be drawn regarding the interval developed for EPCP for humans in Section 6.5.1. As before, when the EPCP was considered, this estimate seems high; however, the representative nature of the data to the general population is unknown.

![U-Shape Curve for Humans](image)

Figure 6: A U-Shape Curve for humans.

**Estimating the Odds at $x_{\text{MIN D,E}}$**

Using equation (4.84), the log odds of an excess condition are estimated by

$$\hat{\beta}_0 + \hat{\beta}_2 + (\hat{\beta}_1 + \hat{\beta}_3)x_{\text{MIN D,E}} = -9.513 + 15.95 + (-4.655 + 11.58)(-1.408) = -3.319,$$

which suggests the odds are $e^{-3.319} = 0.036$. There is a $\frac{3.6}{96.4} = 3.73\%$ chance that a dose of 0.039 mg/kg bw will lead to severe health effects.

Using equation (4.83), the log odds of a deficiency condition are estimated by

$$\hat{\beta}_0 + \hat{\beta}_1x_{\text{MIN D,E}} = -9.513 + (-4.655)(-1.408) = -2.96,$$

which suggests the odds are $e^{-2.96} = 0.052$. There is a $\frac{5.2}{94.8} = 5.5\%$ chance that a dose of 0.039 mg/kg bw will lead to severe health effects. Again, this estimate seems
high. When a dose of 0.039 mg/kg bw is administered, the odds suggest a non-normal reading attributed to a deficient condition is more likely.

6.7.2 Mice

**Estimating $x_{\text{MIN \ D\&E}}$**

A plot of $P(\text{SEV})$ versus log concentration is given in Figure 7. For mice, $x_{\text{MIN \ D\&E}}$ is estimated by 0.209 using the Newton Raphson approach described in Section 4.11. A dose of $10^{0.209} = 1.6183$ mg/kg bw will minimize the probability of a non-normal reading emerging from an excess or deficient condition, or both. In addition, using the methodology described in Section 5.3, we can construct a 95% bootstrap confidence interval for $x_{\text{MIN \ D\&E}}$. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, $10^{x_{\text{MIN \ D\&E}}}$ lies between 1.50978 and 2.0729 mg/kg bw.

![U-Shape Curve for Mice](image)

Figure 7: A U-Shape Curve for mice.
Estimating the Odds at $x_{\text{MIN DUE}}$

Using equation (4.86), the log odds of an excess condition are estimated by 
$$\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 + (\hat{\beta}_1 + \hat{\beta}_5 + \hat{\beta}_6) x_{\text{MIN DUE}} = -9.513 + 15.95 + 6.436742 - 17.46571 + (-4.655 + 11.58 - 4.893107)(0.209) = -4.167,$$ 
which suggests the odds are $e^{-4.167} = 0.0155$. There is a $\frac{0.55}{0.45} = 1.23\%$ chance that a dose of 1.62 mg/kg bw will lead to severe health effects.

Using equation (4.85), the log odds of a deficient condition are estimated by 
$$\hat{\beta}_0 + \hat{\beta}_3 + (\hat{\beta}_1 + \hat{\beta}_6) x_{\text{MIN DUE}} = -9.513 + 6.436742 + (-4.655 - 4.893107)(0.209) = -5.07,$$ 
which suggests the odds are $e^{-5.07} = 0.0063$. 
There is a $\frac{0.63}{0.37} = 1.63\%$ chance that a dose of 1.62 mg/kg bw will lead to severe health effects. When a dose of 1.16183 mg/kg bw is administered, the odds suggest a non-normal reading attributed to an excess condition is more likely.

6.7.3 Rats

Estimating $x_{\text{MIN DUE}}$

A plot of $P(\text{SEV})$ versus log concentration is given in Figure 8. For rats, $x_{\text{MIN DUE}}$ is estimated by -0.162 using the Newton Raphson approach described in Section 4.11. A dose of $10^{-0.162} = 0.688$ mg/kg bw will minimize the probability of a non-normal reading attributed to an excess or deficiency condition, or both. In addition, using the methodology described in Section 5.3, we can construct a 95% bootstrap confidence interval for $x_{\text{MIN DUE}}$. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, $10^{x_{\text{MIN DUE}}}$ lies between 0.5745 and 0.8113.
mg/kg bw.

Figure 8: A U-Shape Curve for rats.

Estimating the Odds at $x_{\text{MIN DUE}}$

Using equation (4.88), the log odds of an excess condition are estimated by

$$\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_4 + \hat{\beta}_9 + (\hat{\beta}_1 + \hat{\beta}_5 + \hat{\beta}_7)x_{\text{MIN DUE}} = -9.513 + 15.95 + 3.4165 - 13.158 + (-4.655 + 11.58 - 4.833)(-0.162) = -3.643,$$

which suggests the odds are $e^{-3.643}=0.026$. There is a $\frac{26}{100}=2.6\%$ that a dose of 0.688 mg/kg bw will lead to severe health effects.

Using equation (4.87), the log odds of a deficiency condition are estimated by

$$\hat{\beta}_0 + \hat{\beta}_4 + (\hat{\beta}_1 + \hat{\beta}_7)x_{\text{MIN DUE}} = -9.513 + 3.4165 + (-4.655 - 4.833)(-0.162) = -4.559,$$

which suggests the odds are $e^{-4.559}=0.0105$. There is a $\frac{105}{985}=1.06\%$ chance that a dose of 0.688 mg/kg bw will lead to severe health effects. When a dose of 0.688 mg/kg bw is administered, the odds suggest a non-normal reading attributed to an excess condition is more likely.
7 Application with Manganese Database

In this chapter, we present the results obtained when we simultaneously model excess and deficiency conditions using a manganese(Mn) database. This second application is intended to be another example illustrating the methodology developed for this thesis.

The same modifications applied to the copper database were also be applied to the Mn database. These include:

- conversion of group-level data into individual-level data
- ordinal response variable collapsed to binary response variable
- log concentration as main effect with extensions to include analyses with stratification by species

Note that the same limitations surrounding the use of EPCP and $x_{MIN \cup E}$ when the copper database was considered are applicable here. Specifically, we consider models with and without species stratification. The latter, presented in Sections 7.1 through 7.4, and 7.6, is simply included as an illustration of the suggested approaches, since species confounding will compromise the ability to draw any health-related conclusions.

7.1 Representations of the JMED

Fitting the model in (4.1) to the available data in the Mn database yields $\hat{\beta}_0 = -1.0464, \hat{\beta}_1 = -2.6875, \hat{\beta}_2 = -1.381,$ and $\hat{\beta}_3 = 4.3543,$ along with an associated estimated covariance matrix. A plot of the excess and deficiency curves resulting from this fit is given in Figure 9.
7.2 Interpreting the EPCP

The EPCP is estimated by $-\frac{\hat{\beta}_2}{\hat{\beta}_3} = \frac{1.381}{4.354} = 0.31718$. When a subject ingests $10^{0.31718} = 2.076$ mg/kg bw, the probability of experiencing adverse health effects from excess exposure is equivalent to the probability of experiencing adverse health effects from deficient exposure. A daily exposure of 2.076 mg/kg bw will protect an individual from a harmful ‘unknown’ Mn condition. In addition, using equation (4.10), we can compute a 95% confidence interval for EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, $10^{\text{EPCP}}$ lies between 1.7407 and 2.475 mg/kg bw.
7.3 Interpreting the Odds of Excess or Deficiency at the EPCP

Using equation (4.12), the log odds of severity at the EPCP are estimated as 
\[ \hat{\beta}_0 - \frac{\hat{\beta}_1 \hat{\beta}_2}{\hat{\beta}_3} = -1.0464 + \frac{-2.6875(-1.381)}{4.354} = -1.8988, \]
which suggests that the odds of severity are estimated by \( e^{-1.89876} = 0.14975 \). At the EPCP, we estimate a severe health effect will occur in \( \frac{14.98}{100-14.98} = 17.62\% \) of all cases. At the EPCP, the odds are in favour of not experiencing a severe health effect. In addition, using equation (4.18), we can compute a 95% confidence interval for the log odds at EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, the odds lies between 0.1293 and 0.1735.

7.4 Stratification by Species

To perform a regression analysis focused on stratification by species, three new subsets have been created comprised of observations on mice, rats, and all remaining species, which will be referred to as ‘others.’ The ‘others’ category includes humans, monkeys, guinea pigs, hamsters, and pigs. The excess and deficiency data sets of the ‘others’ are distributed differently. For example, there are no hamsters in the deficiency data set. Similar to the Cu database, the primary focus of this application is to interpret the results in the context of toxicity and health.

Fitting the JMED model in (4.24) to the mice, rats, and ‘others’ data described above yields \( \hat{\beta}_0 = -3.142914, \hat{\beta}_1 = -4.533164, \hat{\beta}_2 = 2.925715, \hat{\beta}_3 = 2.224124, \hat{\beta}_4 = 3.40254, \hat{\beta}_5 = 4.751321, \hat{\beta}_6 = 1.837297, \hat{\beta}_7 = 2.689746, \)

107
\( \hat{\beta}_8 = -4.711887 \), and \( \hat{\beta}_9 = -8.572167 \), along with an associated estimated covariance matrix. A plot of the excess and deficiency curves resulting from this fit is given in Figure 10.

![Excess and Deficiency Curves](image)

**Figure 10:** Species stratified models for Mn Database.

### 7.4.1 Interpreting EPCP\(_{\text{OTH}}\), EPCP\(_{\text{MU}}\), EPCP\(_{\text{RT}}\)

**Others**

The EPCP is estimated by 
\[
-\frac{\hat{\beta}_9}{\hat{\beta}_8} = \frac{-2.925}{4.751} = -0.6158
\]

For others, at a daily exposure concentration of \(10^{-0.6158} = 0.2455\) mg/kg bw, the probability of experiencing adverse health effects from excess exposure is equivalent to the probability of experiencing adverse health effects from deficient exposure. In addition, using equation (4.10), we can compute a 95% confidence interval for EPCP\(_{\text{OTH}}\). Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, \(10^{EPCP_{\text{OTH}}}\) lies between
0.1878 and 0.3124 mg/kg bw.

**Mice**

The EPCP is estimated by $-\hat{\beta}_2 + \hat{\beta}_8 = -\frac{2.925-4.712}{4.751} = 0.3759$. For mice, at a daily exposure concentration of $10^{0.3759} = 2.38$ mg/kg bw, the probability of experiencing adverse health effects from excess exposure is equivalent to the probability of experiencing adverse health effects from deficient exposure. In addition, using equation (4.30), we can compute a 95% confidence interval for $EPCP_{MU}$. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, $10^{EPCP_{MU}}$ lies between 1.6488 and 3.425 mg/kg bw.

**Rats**

The EPCP is estimated by $-\hat{\beta}_2 + \hat{\beta}_9 = -\frac{2.925-8.5721}{4.751} = 1.188$. For rats, at a daily exposure concentration of $10^{1.1884} = 15.43$ mg/kg bw, the probability of experiencing adverse health effects from excess exposure is equivalent to the probability of experiencing adverse health effects from deficient exposure. In addition, using equation (4.30) we can compute a 95% confidence interval for $EPCP_{RT}$. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, $10^{EPCP_{RT}}$ lies between 8.039 and 29.621 mg/kg bw.
7.4.2 Interpreting the Log Odds at the EPCP

Others

Using equation (4.12), the log odds of severity at the EPCP are estimated as

\[ \hat{\beta}_0 - \frac{\hat{\beta}_1 \hat{\beta}_2}{\hat{\beta}_5} = -3.143 - \frac{(-4.533)(2.926)}{4.751} = -0.3515, \]

which suggests the odds are estimated by \( e^{-0.3515} = 0.7036 \). For others, the odds strongly favour a severe health effect when a dose of 0.03905 mg/kg bw is administered. In addition, using equation (4.18), with \( C^T \) defined in (4.33), we can compute a 95% confidence interval for the log odds at the EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, the odds lie between 0.2419 and 2.046.

Mice

Using equation (4.34), the log odds of severity at the EPCP are estimated as

\[ \hat{\beta}_0 + \frac{\hat{\beta}_3}{\hat{\beta}_5} = -3.143 + 2.224 - \frac{(-4.533+1.837)(2.926-4.711)}{4.751} = -1.932, \]

which suggests the odds are estimated by \( e^{-1.932} = 0.1448 \). For a mouse, there is a \( \frac{14.48}{85.32} = 16.93\% \) chance of experiencing a severe health effect. At the EPCP, the odds are in favour of not experiencing a severe health effect. In addition, using equation (4.41), we can compute a 95% confidence interval for the log odds at the EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, the odds lie between 0.1252 and 0.1675.
Rats

Using equation (4.35), the log odds of severity at the EPCP are estimated as
\[ \hat{\beta}_0 + \hat{\beta}_4 - \frac{(\hat{\beta}_1 + \hat{\beta}_7)(\hat{\beta}_2 + \hat{\beta}_9)}{\hat{\beta}_5} = -3.143 + 3.402 - \frac{(-4.533 + 2.689)(-8.926 - 8.572)}{4.751} = -1.937, \]
which suggests the odds are estimated by \( e^{-1.937} = 0.1441 \). For a rat, there is a \( \frac{14.41}{85.59} = 16.84\% \) chance of experiencing a severe health effect when a dose of 0.5742 mg/kg bw is administered. In addition, using equation (4.41), we can compute a 95% confidence interval for the log odds at the EPCP. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, the odds lie between 0.1074 and 0.1957.

7.4.3 Species-to-Species Odds Ratios at the EPCP

Mice to Others

Using equation (4.45), the odds ratio for mice to others is estimated as
\[ e^{\hat{\beta}_4 - \frac{\hat{\beta}_2 \hat{\beta}_6 + \hat{\beta}_1 \hat{\beta}_9 + \hat{\beta}_6 \hat{\beta}_9}{\hat{\beta}_5}} = 0.2058 \]
At doses of 2.38 mg/kg bw and 0.245 mg/kg bw for mice and others, respectively, mice are 0.2058 times more likely to exhibit a non-normal condition. Therefore, others are more sensitive to a non-normal condition than mice. In addition, using equation (4.51), we can compute a 95% confidence interval for the odds ratio. With 95% confidence, we estimate the odds ratio lies between 0.1014 and 0.4179.

Rats to Others

Using equation (4.46), the odds ratio for rats to humans is estimated as
\[ e^{\hat{\beta}_4 - \frac{\hat{\beta}_2 \hat{\beta}_6 + \hat{\beta}_1 \hat{\beta}_9 + \hat{\beta}_6 \hat{\beta}_9}{\hat{\beta}_5}} = 0.2061 \]
At doses of 15.43 mg/kg bw and 0.245 mg/kg bw for rats and others, respectively, rats are 0.2061 times more likely to exhibit a
non-normal condition. Therefore, others are more sensitive to a non-normal condition than rats. In addition, using equation (4.51), we can compute a 95% confidence interval for the odds ratio. With 95% confidence, we estimate the odds ratio lies between 0.0907 and 0.4678.

**Rats to Mice**

Using equation (4.52), the odds ratio for rats to mice is estimated as

\[
e^{\hat{\beta}_4-\hat{\beta}_3-(\hat{\beta}_2+\hat{\beta}_9)(\hat{\beta}_1+\hat{\beta}_7)+(\hat{\beta}_2+\hat{\beta}_8)(\hat{\beta}_1+\hat{\beta}_6)}
\]

\[= 0.139.\]

At doses of 15.43 mg/kg bw and 2.38 mg/kg bw for rats and mice, respectively, rats are 0.139 times more likely to exhibit a non-normal condition. In addition, using equation (4.58), we can compute a 95% confidence interval for the odds ratio. With 95% confidence, we estimate the odds ratio lies between 0.078 and 0.2218.

### 7.5 An Alternative Approach: \( x_{\text{MINDUE}} \)

Using the Newton Raphson approach described in Section 4.11, \( x_{\text{MINDUE}} \) is estimated by 0.3269. For all species, a concentration of \( 10^{0.3269} = 2.12 \) mg/kg bw will minimize the probability of severity emerging from an excess condition, a deficient condition, or both. In addition, using the methodology described in Section 5.3, we can construct a 95% bootstrap confidence interval for \( x_{\text{MINDUE}} \). Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, \( 10^{x_{\text{MINDUE}}} \) lies between 2.0577 and 2.3302 mg/kg bw.
7.5.1 Interpreting the Odds of Excess or Deficiency at $x_{\text{MIN,DJE}}$

Using equation (4.66), the log odds of a deficiency condition are estimated by $\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{\text{MIN,DJE}} = -1.0464 + (-2.6875)(0.3269) = -1.925$, which suggests an estimate for the odds is $e^{-1.925} = 0.1459$. Assuming a deficiency condition, there is a $\frac{14.59}{85.41} = 17.08\%$ chance that a dose of 2.12 mg/kg bw will lead to a severe health effect.

Using equation (4.67), the log odds of an excess condition are estimated by $\hat{\beta}_0 + \hat{\beta}_2 + (\hat{\beta}_1 + \hat{\beta}_3) \hat{x}_{\text{MIN,DJE}} = -1.0464 - 1.381 + (-2.6875 + 4.354)(0.3269) = -1.883$, which suggests an estimate for the odds is $e^{-1.883} = 0.1521$. Assuming a deficiency condition, there is a $\frac{15.21}{84.79} = 17.94\%$ chance that a dose of 2.12 mg/kg bw/day will lead to a severe health effect. At $x_{\text{MIN,DJE}}$, the odds suggest a non-normal reading caused by an excess condition is more likely.
7.6 The Independence Model Applied to the Species-Stratified JMED

Similar to Section 6.7, using the fit of the JMED in (4.24), we computed a point estimate and bootstrap confidence interval for each species-specific \( x_{\text{MIN} \cup \text{E}} \), along with estimates of the odds of deficient and excess conditions at this log concentration. We present plots of \( P(\text{SEV}) \) for each species, given in (4.70), (4.75), and (4.80).

7.6.1 Others

**Estimating \( x_{\text{MIN} \cup \text{E}} \)**

A plot of \( P(\text{SEV}) \) versus log concentration is given in Figure 12. For others, \( x_{\text{MIN} \cup \text{E}} \) is estimated by -0.406 using the Newton Raphson approach described in Section 4.11. A dose of \( 10^{-0.406} = 0.39237 \text{ mg/kg bw} \) will minimize the probability of a non-normal reading emerging from an excess condition, a deficient condition, or both. In addition, using the methodology described in Section 5.3, we can construct a 95% bootstrap confidence interval for \( x_{\text{MIN} \cup \text{E}} \). Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, we estimate that \( 10^{x_{\text{MIN} \cup \text{E}}} \) lies between 0.3913 and 0.4225 mg/kg bw.

**The Odds at \( x_{\text{MIN} \cup \text{E}} \)**

Using equation (4.84), the log odds of an excess condition are estimated by \( \hat{\beta}_0 + \hat{\beta}_2 + (\hat{\beta}_1 + \hat{\beta}_5)\hat{x}_{\text{MIN} \cup \text{E}} = -0.30577 \), which suggests the odds can be estimated as \( e^{0.30577} = 0.736 \). The odds strongly favour a severe excess

114
condition when a dose of 0.3924 mg/kg bw is administered.

Using equation (4.83), the log odds of a deficiency condition are estimated
by $\hat{\beta}_0 + \hat{\beta}_1 x_{\text{MIN DxE}} = -1.302$, which suggests the odds can be estimated as $e^{-1.302} = 0.272$. There is a $\frac{.474}{1.28} = 37.36\%$ chance that a dose of 0.3924 mg/kg bw will lead to severe health effects. For others, the odds suggest an excess condition is more likely to cause adverse health effects.

7.6.2 Mice

Estimating $x_{\text{MIN DxE}}$

A plot of $P(\text{SEV})$ versus log concentration is given in Figure 13. For mice, $x_{\text{MIN DxE}}$ is estimated to be 0.3759 using the Newton Raphson approach described in Section 4.11. A dose of $10^{0.3759} = 2.3763$ mg/kg bw will minimize the likelihood of a non-normal reading emerging from an excess condition, a
deficient condition, or both. In addition, using the methodology described in Section 5.3, we can construct a 95% bootstrap confidence interval for $x_{\text{MIN DUE}}$. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, that $10^{x_{\text{MIN DUE}}}$ lies between 2.218 and 2.7713 mg/kg bw.

![U-Shape Curve for Mice](image)

**Figure 13: A U-Shape Curve for mice.**

**The Odds at $x_{\text{MIN DUE}}$**

Using equation (4.86), the log odds of an excess condition are estimated by

$$\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_6 + (\hat{\beta}_1 + \hat{\beta}_5 + \hat{\beta}_6) x_{\text{MIN DUE}} = -1.932,$$

which suggests the odds can be estimated as $e^{-1.932} = 0.1448$. There is a $\frac{14.48}{85.52} = 16.93\%$ chance of a severe excess condition when a dose of 2.377 mg/kg bw is administered. Using equation (4.85), the log odds of a deficiency condition are estimated by

$$\hat{\beta}_0 + \hat{\beta}_3 + (\hat{\beta}_1 + \hat{\beta}_6) x_{\text{MIN DUE}} = -1.932,$$

which suggests the odds can be estimated
as $e^{-1.932} = 0.1448$. There is a $\frac{14.48}{85.52} = 16.93\%$ chance of a severe deficiency condition when a dose of 2.377 mg/kg bw is administered. Coincidentally, the odds of a severe condition are the same for excess and deficiency conditions.

### 7.6.3 Rats

**Estimating $x_{MIN\ DU:E}$**

A plot of $P(\text{SEV})$ versus log concentration is given in Figure 14. For rats, $x_{MIN\ DU:E}$ is estimated to be 1.097. A dose of $10^{1.097} = 12.531$ mg/kg bw will minimize the probability of severity emerging from an excess condition, a deficient condition, or both. In addition, using the methodology described in Section 5.3, we can construct a 95% bootstrap confidence interval for $x_{MIN\ DU:E}$. Appropriately transforming the upper and lower limits of this interval allows us to conclude with 95% confidence, that $10^{x_{MIN\ DU:E}}$ lies between 10.2093 and 17.9887 mg/kg bw.

![U-Shape Curve for Rats](image)

**Figure 14:** A U-Shape Curve for rats.
The Odds at \( x_{\text{MIN D} \cup E} \)

Using equation (4.88), the log odds of an excess condition are estimated by
\[
\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_4 + \hat{\beta}_0 + (\hat{\beta}_1 + \hat{\beta}_5 + \hat{\beta}_7)\hat{x}_{\text{MIN D} \cup E} = -2.197,
\]
which suggests the odds can be estimated as \( e^{-2.197} = 0.1111 \). There is a \( \frac{1111}{8889} = 12.50\% \) chance of a severe excess condition when a dose of 12.531 mg/kg bw is administered.

Using equation (4.87), the log odds of a deficiency condition are estimated by
\[
\hat{\beta}_0 + \hat{\beta}_4 + (\hat{\beta}_1 + \hat{\beta}_7)\hat{x}_{\text{MIN D} \cup E} = -1.763,
\]
which suggests the odds can be estimated as \( e^{-1.763} = 0.1715 \). There is a \( \frac{1715}{8285} = 20.7\% \) chance of a severe deficiency condition when a dose of 12.531 mg/kg bw is administered.
8 Conclusion

The simultaneous modeling of excess and deficiency conditions to characterize U-shaped exposure-response relationships was the central focus of this project. The Independence Model achieves this objective. The Joint Model for Excess and Deficiency (JMED), expressed as a well-defined statistical model, provided the foundation. In fitting these models to the copper database studied by Chambers et al. (2010), the U-shape curve that results from the Independence Model when the probability of an adverse condition is plotted against log concentration is asymmetrical, reflecting the differential tendencies exhibited by Cu excess and deficient conditions. The estimation of the bottom of the U, formally stated as $x_{MIND∪E}$, surfaces as a minimization problem requiring numerical methods. Note that it is also possible to compute a confidence interval for $x_{MIND∪E}$ using bootstrap techniques.

The JMED model gave rise to a second point of interest with useful properties. The *equiprobable crossover point* (EPCP) is defined as the point on the horizontal axis of a plot of the probability of an adverse condition versus log concentration where the excess and deficiency curves cross. This point differs from $x_{MIND∪E}$ because it represents the exposure concentration that ‘protects’ an individual from an unknown condition that has the potential to inflict health problems. A closed form solution for the EPCP can be obtained. The closed form solution facilitates the development of approximate expressions for the variance and bias of the estimator using Taylor series expansions. This is valuable because confidence intervals can be constructed without resorting to bootstrap techniques, although intervals based on the
latter approach are also possible. Both types of intervals perform admirably well in terms of coverage provided the denominator of the expression for EPCP is not close to zero. For the JMED and Independence Models that do not stratify by species, care must be taken in interpreting the associated results. For example, drastic differences in estimates of $x_{\text{MIND}} \cup E$ and EPCP are likely to be observed for different species.

The inclusion of mice and rat indicator variables facilitated a species-specific analysis of the JMED model. The EPCP estimate for humans is 0.042 mg/kg bw. At a daily exposure of 0.042 mg/kg bw, the probability of a non-normal health condition is approximately 4.7%. Among humans, mice and rats, humans have the highest chance of a non-normal reading at the EPCP, while mice have the lowest chance. Applying the Independence Model to the Species-Stratified JMED model provides an alternative approach.

The $x_{\text{MIND}} \cup E$ for humans is 0.039 mg/kg bw. Assuming a deficiency condition, there is a 5.5% chance of a severe health effect. Conversely, there is a 3.73% chance of a severe health effect from an excess condition. At $x_{\text{MIND}} \cup E$, a deficiency condition puts humans at a higher risk of a non-normal reading than an excess condition. In terms of prescribing health recommendations, the chance of a severe health effect attributed to either condition seems fairly high. It is very unlikely that the proportion of those exhibiting a severe condition is representative of the entire population, suggesting our estimation may be inflated. Among humans, mice and rats, humans have the highest chance of a non-normal reading at $x_{\text{MIND}} \cup E$, while mice have the lowest chance.

A discernible difference exists in the estimates for the EPCP and $x_{\text{MIND}} \cup E$. 
for the general and species-specific models. Whereas $x_{\text{MIND} \cup \text{E}}$ is more exhaustive because it ensures the subject is protected from adverse health effects attributed to an excess condition, a deficient condition, or both, the EPCP protects an individual from venturing into a range of exposure concentrations considered severely excess or severely deficient. Based on this interpretation, $x_{\text{MIND} \cup \text{E}}$ is the most appropriate recommendation for the single optimal intake level. Assuming bodyweight is 70 kg, the single optimal intake level for humans emerges as 2.73 mg Cu/day. It is logical to expect that there is an acceptable range for copper intake, not just a single value. We can use the 95% bootstrap confidence interval to define this range. Our study results suggest an acceptable range of oral intake between 1.577 and 4.462 mg Cu/day. This estimation is close to previous work completed by Chambers et al. (2010), where the single optimal intake level for humans was estimated to be 2.6 mg Cu/day.

Chambers et al. (2010) used CatReg to conduct an exposure-response analyses on the copper database. Excess and deficiency conditions had to be modeled separately. The excess observations were best described by an ordinal regression model, while the deficiency observations were represented by a binary logistic regression model. Stratification factors such as age and sex were applied differently to the excess and deficiency models. In summary, the excess and deficiency models used to represent the copper database differ in terms of their response variable and stratification factors and are ultimately impossible to model simultaneously. In order to fit excess and deficiency curves simultaneously, it became necessary to find an alternative approach to CatReg.
Using binary logistic regression as a foundation, the JMED was developed to predict the probability of a non-normal reading with knowledge of study classification (excess or deficiency) and log concentration, and most importantly, facilitates the simultaneous fitting of excess and deficiency models. To extend the model, future initiatives could focus on modeling with an ordinal response variable. Finally, the data available on humans is limited due to the ethical considerations involved in administering Cu in excess or deficient amounts. Developing a model to extrapolate from rats to humans would resolve the issues surrounding the limited amounts of data available for humans.
References


123


