



---

Theses and Dissertations

---

2007-10-25

## Applying Bayesian Ordinal Regression to ICAP Maladaptive Behavior Subscales

Edward P. Johnson  
*Brigham Young University - Provo*

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Statistics and Probability Commons](#)

---

### BYU ScholarsArchive Citation

Johnson, Edward P., "Applying Bayesian Ordinal Regression to ICAP Maladaptive Behavior Subscales" (2007). *Theses and Dissertations*. 1210.  
<https://scholarsarchive.byu.edu/etd/1210>

This Selected Project is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

APPLYING BAYESIAN ORDINAL REGRESSION TO ICAP  
MALADAPTIVE BEHAVIOR SUBSCALES

by

Edward Paul Johnson

A project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

December 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Edward Paul Johnson

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

\_\_\_\_\_

Date

\_\_\_\_\_

Gilbert W. Fellingham, Chair

\_\_\_\_\_

Date

\_\_\_\_\_

Scott D. Grimshaw

\_\_\_\_\_

Date

\_\_\_\_\_

David G. Whiting

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Edward Paul Johnson in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

\_\_\_\_\_  
Date

\_\_\_\_\_  
Gilbert W. Fellingham  
Chair, Graduate Committee

Accepted for the Department

\_\_\_\_\_  
Date

\_\_\_\_\_  
Scott D. Grimshaw

Accepted for the College

\_\_\_\_\_  
Date

\_\_\_\_\_  
Tom Sederberg, Associate Dean  
College of Physical & Mathematical Sciences

## ABSTRACT

### APPLYING BAYESIAN ORDINAL REGRESSION TO ICAP

### MALADAPTIVE BEHAVIOR SUBSCALES

Edward Paul Johnson

Department of Statistics

Master of Science

This paper develops a Bayesian ordinal regression model for the maladaptive subscales of the Inventory for Client and Agency Planning (ICAP). Because the maladaptive behavior section of the ICAP contains ordinal data, current analysis strategies combine all the subscales into three indices, making the data more interval in nature. Regular MANOVA tools are subsequently used to create a regression model for these indices. This paper uses ordinal regression to analyze each original scale separately. The sample consists of applicants for aid from Utah's Division of Services for Persons with Disabilities. Each applicant fills out the *Scales of Independent Behavior—Revised* (SIB-R) portion of the ICAP that measures eight different maladaptive behaviors. This project models the frequency and severity of each of these eight problem behaviors with separate ordinal regression models. Gender, ethnicity, primary disability, and mental

retardation are used as explanatory variables to calculate the odds ratios for a higher maladaptive behavior score in each model. This type of analysis provides a useful tool to any researcher using the ICAP to measure maladaptive behavior.

*Key words:* latent variable, SIB-R, disabilities research, proportional odds model



## TABLE OF CONTENTS

CHAPTER	
1	Introduction 1
1.1	Introduction to ICAP and SIB-R.....1
1.2	Introduction to Ordinal Regression.....3
2	Literature Review 5
2.1	Development of Ordinal Regression.....5
2.2	Johnson's Bayesian Hierarchal Ordinal Regression Model.....6
2.3	Possible Prior Distributions for a Bayesian Ordinal Regression Model.....11
2.4	MCMC and Cowles' Hybrid Metropolis-Hastings Algorithm .....13
3	Methodology 16
3.1	Formulation of the Bayesian Ordinal Regression Model .....16
3.2	Simulation from the Complete Conditional Distributions.....18
4	Results 21
4.1	Validity of the Frequency Models .....21
4.2	Validity of the Severity Models.....28
4.3	Results of the Bayesian Ordinal Regression Models.....35
5	Conclusions 40
BIBLIOGRAPHY 42	
APPENDIX	
A	Computer Code 46



## LIST OF TABLES

### TABLE

3.1	Frequency of Categories of Gender .....	16
3.2	Frequency of Categories of Disability .....	17
3.3	Frequency of Categories of Mental Ability .....	17
3.4	Frequency of Categories of Ethnicity .....	17
3.5	Frequency of Categories of Marital Status .....	18
4.1	Credible Intervals for the Parameters in the Frequency Models.....	38
4.2	Credible Intervals for the Parameters in the Severity Models .....	39

## LIST OF FIGURES

### FIGURE

2.1	Distribution of the Latent Variable .....	9
4.1	Mixing Plot for the Intercept.....	21
4.2	Mixing Plots for the Cutoff Parameters .....	22
4.3	Mixing Plot for the Gender Parameter.....	23
4.4	Posterior Distribution of the Odds Ratio of Men to Women .....	24
4.5.	Graph of Responses to Hurtful to Self by Gender .....	24
4.6	Mixing Plot for an Autism Parameter .....	25
4.7	Posterior Distribution of the Odds Ratio of Autism to Mental Retardation .....	25
4.8	Mixing Plot for a Cerebral Palsy Parameter .....	26
4.9	Posterior Distribution of the Odds Ratio of Cerebral Palsy to Mental Retardation.....	26
4.10	Graph of Responses to Hurtful to Self by Disability .....	27
4.11	Mixing Plot for the Intercept.....	28
4.12	Mixing Plots for the Cutoff Parameters .....	29
4.13	Mixing Plot for the Gender Parameter.....	30
4.14	Posterior Distribution of the Odds Ratio of Men to Women .....	30
4.15	Graph of Responses to Hurtful to Self by Gender .....	31
4.16	Mixing Plot for an Autism Parameter .....	32
4.17	Posterior Distribution of the Odds Ratio of Autism to Mental Retardation .....	32
4.18	Mixing Plot for a Cerebral Palsy Parameter .....	33

4.19	Posterior Distribution of the Odds Ratio of Cerebral Palsy to Mental Retardation.....	33
4.20	Graph of Responses to Hurtful to Self by Disability .....	34

## 1. INTRODUCTION

### 1.1 Introduction to ICAP and SIB-R

Families that have the responsibility to care for individuals with disabilities often experience financial challenges associated with those handicaps. These families can obtain governmental support; however, the demand for assistance often outstrips available funding. Consequently, governmental agencies require a testing instrument that quantifies the degree of family need. Governmental agencies use the results of these testing instruments to determine fund allocation. One of the instruments in current use is the *Inventory for Client and Agency Planning* (ICAP) (Bruininks et al. 1986). Utah, like many other states, uses the ICAP to allocate funding.

Many researchers also use the ICAP to measure the effectiveness of disability programs (Thompson et al. 2002; Orsmond et al. 2003). One part of the ICAP, the *Scales of Independent Behavior—Revised* (SIB-R), measures the overall behavior of the disabled individual. A section of the SIB-R addresses the maladaptive behavior of the individual. An interviewer administers the SIB-R by asking a parent or caregiver scripted questions about the disabled dependent. The scripted questions include complete instructions for the interviewer on how to elicit responses with minimal interviewer bias (Bruininks et al. 1996). The SIB-R measures the frequency and severity of eight different problem behaviors using Likert scales. The frequency scales range from 0 to 5 while the severity scales range from 0 to 4.

For example, the interviewer asks, “Does (name) injure his/her own body—for example, by hitting self, banging head, scratching, cutting or puncturing, biting, rubbing skin, pulling out hair, picking on skin, biting nails, or pinching self?” This introduction

gives the parent or caregiver a thorough description of actions associated with the problem behavior. After recording the response, the interviewer asks, “How often does this behavior occur?” and “How serious is the problem usually caused by this behavior?” The respondent then selects a frequency from the following scale: [0—Never, 1—Less than once a month, 2—One to 3 times a month, 3—One to 6 times a week, 4—One to 10 times a day, or 5—One or more times an hour]. Note that the scale does not have consistent time interval, but does progress from least often to most often. The respondent also selects a severity from the following scale: [0—Not serious, not a problem; 1—Slightly serious, a mild problem; 2—Moderately serious, a moderate problem; 3—Very serious, a severe problem; or 4—Extremely serious, a critical problem] (Bruininks et al. 1996). The eight problem behaviors were named “Hurtful to Self,” “Hurtful to Others,” “Destructive to Property,” “Disruptive Behavior,” “Unusual or Repetitive Habits,” “Socially Offensive Behavior,” “Inattentive Behavior,” and “Uncooperative Behavior.” All eight problem behaviors have a corresponding question and procedure.

At the end of the interview, the SIB-R has sixteen measurements (eight severity scores and eight frequency scores). The SIB-R contains an algorithm to translate these sixteen measurements into three different indices: Internalized Maladaptive Behavior, Asocial Maladaptive Behavior, and Externalized Maladaptive Behavior. The SIB-R also creates one grand total scale for an overall summary. Orsmond (2003) and other researchers use these interval scales to identify effects using MANOVA. The translation method combines the different scores for each maladaptive behavior into interval indices, so MANOVA is an appropriate test. However, some researchers need to analyze the sixteen measurements directly, particularly if the treatment affects only one or two of the

problem behaviors. Researchers also are interested in identifying which basic demographics are associated with each problem behavior. Basic demographics included with the SIB-R are marital status, ethnicity, gender, and mental retardation level.

## 1.2 Introduction to Ordinal Regression

Ordinal regression allows researchers to directly measure demographic traits on a single maladaptive behavior scale regardless of continuity. Ordinary least squares regression assumes a normally distributed response variable given the explanatory variables. Applying ordinary least squares regression to Likert scales violates this assumption. Ordinal regression solves this problem by including a continuous latent variable that translates into the observed ordinal scale. Cutoff parameters divide the latent variable's distribution into a discrete number of areas. These areas correspond to the possible responses on the Likert scale. The ability to implement ordinal regression techniques is important any time a Likert scale is used.

Johnson (2003) completed a study using ordinal regression entitled "Duke Undergraduates Evaluate Teaching" (DUET). The study isolated the effect of grades on teacher evaluations. Teacher evaluations were done using multiple Likert scales which evaluated different aspects of teaching. Johnson created a Bayesian hierarchical model with an experimental unit of each individual in each class. He used ordinal regression because the teacher evaluations were not on a continuous scale. He estimated cutoff values to place the continuous latent distribution onto the observable discrete response variable. He found that a student who expected an A was three times more likely to rate a teacher higher on some items than a student who expected a B (Johnson 2003).

This project adapts Johnson's ordinal regression Bayesian hierarchical model to analyze maladaptive behavior scores. Utah's Division of Services for Persons with Disabilities provided the data for over 7,000 individuals. Each respondent completed the SIB-R to qualify for government financial support for disabled persons. This project examines the effect of gender, ethnicity, disability type, and marital status on each of the eight maladaptive behaviors and summarizes which characteristics are associated with higher maladaptive behavior scores.

## 2. LITERATURE REVIEW

### 2.1 Development of Ordinal Regression

McCullagh (1980) developed a technique to analyze ordinal data called ordinal regression. Many times a response variable is not measured on a continuous scale. When this is the case, it is not appropriate to use ordinary least squares regression with ordinal data. McCullagh formulated a way to use regression techniques on a discrete response scale. He modeled the mean of an unobservable continuous latent variable with a linear function of the explanatory variables. McCullagh defined a latent variable to be a characteristic of the subject which is assumed to have a continuous distribution, but is not directly observable; however, the latent variable can be indirectly measured because when the characteristic increases, the probability of survival also increases.

An animal living through winter exemplifies a literal example of this survival concept. Other latent variables use the term “surviving” figuratively, such as a student surviving a class by receiving a passing grade. In the first scenario, the characteristic is the hardiness of the animal. Although not directly measurable, the animal’s hardiness relates to the probability that the animal will live through the winter. If an animal has a high hardiness, then the probability of surviving through winter is quite good. An animal with a low hardiness exhibits a low probability of surviving. Likewise, the latent variable in the second scenario is the “level of preparedness.” These latent variables exist regardless of the ability to measure them.

McCullagh divided the scale of the latent variable into categories with discrete probabilities derived from the areas under the latent variable’s density curve. Survival and death are the categories in the animal example. The categories in the grade example



fall out naturally as A, B, C, D, and F. McCullagh also formulated a more general ordinal regression model that did not assume the existence of a latent variable. He created the distribution of the response variable using the cumulative distribution of the odds ratio. McCullagh also demonstrated other properties of ordinal regression, such as the accuracy of least squares and maximum likelihood estimates.

The ordinal regression model has evolved over time. Albert and Chib (1993) applied the Bayesian framework to the ordinal regression model after the development of Markov Chain Monte Carlo (MCMC) and Gibbs sampling and returned to the latent variable concept. They compared Bayesian and maximum likelihood estimators and found that when the number of observations grew, the estimators converged in distribution. Johnson (1996) used this model in the field of education for essay-rating by teacher assistants. Bedrick et al. (1996) generalized this model by including informative prior distributions on the coefficients for the effects and on the cutoff values for the latent variable using a conditional means approach. Ishwaran and Gatsonis (2000) then expanded the flexibility of the ordinal regression model by including a more complex correlation structure in the prior distribution of the latent variable.

## 2.2 Johnson's Bayesian Hierarchical Ordinal Regression Model

Johnson (2003) wanted to find a way to measure the effect of grade inflation on teacher evaluations. He believes that students rewarded professors who gave them better grades with a better evaluation. Because professors want higher evaluations, which play an increasingly important role in determining tenure and salary, Johnson believes that professors realize they could achieve higher teacher evaluations by lowering their grading standards. Johnson also believes that the resulting grade inflation undermines true

learning. He collected data to study this phenomenon at Duke University in a study called Duke Undergraduates Evaluate Teaching (DUET).

Johnson applied Bayesian ordinal regression in his DUET study, specifying the experience of a student in a class to be his latent variable. This experience ranged from extremely negative to extremely positive. If a student had an extremely positive experience in the class, then the teacher rating would be higher. Johnson used a logistic density function for the likelihood of the latent variable (Johnson 2002, 1999). The general form of the logistic distribution with a location parameter  $\mu$  and a scale parameter  $\sigma$  is

$$f(z; \mu, \sigma) = \frac{e^{-(z-\mu)/\sigma}}{\sigma [1 + e^{-(z-\mu)/\sigma}]^2}. \quad (2.1)$$

Random effects for the course and the student allowed the location parameter  $\mu$  of the latent variable  $z$  to vary for each student in each course. The fixed grade effect enabled the model to measure the relative amount of grade inflation. Ten categories of expected grades (D/F, C-, C, C+, B-, B, B+, A-, A, and A+) spanned the sample space. Johnson used  $X$  to represent the matrix containing  $K$  indicator variables, where  $K$  represented the number of possible expected grade categories. He used  $W$  to represent the matrix containing  $J$  (the number of possible courses) indicator variables. This formulation created the following equation for the latent variable:

$$z_{ijl} = \sum_{k=1}^K x'_{kl} \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w'_{ijl} u_{ijl} + e_{ijl}, \quad (2.2)$$

where  $z_{ijl}$  denotes the latent variable for student  $i$  in course  $j$  on item  $l$ . The fixed effect,  $\beta_{kl}$ , measures how the student's grade affected item  $l$  on the evaluation. The random course effect,  $u_{ijl}$ , varied for every student  $i$  in course  $j$  for item  $l$  of the evaluation. These

column vectors of parameters are multiplied by the respective row vectors found in the  $X$  and  $W$  matrices. These row vectors ( $x'$  and  $w'$ ) are the indicator variables that stem from  $z_{ijl}$ . The error variable,  $e_{ijl}$ , was assumed to be distributed logistically with parameters  $\mu_e=0$  and  $\sigma_e=1$ .

Johnson used this model to analyze data gathered before and after grade distribution. The fixed effects in the first analysis represent the effect of the expected grade. The fixed effects in the second analysis represent the effect of the actual grade received. The random effects remained the same across both models because the course and students remained constant. Johnson estimated cutoff values,  $r$ , for each rating, which translated the latent variable into the ordinal scale of the response variable  $y$ . He scaled the latent variable and addressed identifiability problems by fixing three cutoff values ( $r_0 = -\infty$ ,  $r_1 = 0$ , and  $r_5 = \infty$ ). If the latent variable value lay between  $r_{m-1}$  and  $r_m$ , then the response value was  $m$ .

Figure 2.1 shows the density of a potential latent variable with location parameter of 1 and a spread parameter of 1. Johnson's latent variable distribution translated into multinomial probabilities with  $p_{ijlm}$  denoting the probability that student  $i$  rated course  $j$  on item  $l$  with value  $m$ . The area of the latent variable curve between the two cutoff values equals  $p_{ijlm}$ . The probability distribution in Figure 2.1 shows all possible values of the latent variable. If the latent variable resides between  $r_{m-1}$  and  $r_m$ , then the student gave a rating of  $m$ . The probability of a rating with value  $m$  is equal to the probability of generating a latent variable between  $r_{m-1}$  and  $r_m$  or the integral of the distribution from  $r_{m-1}$  to  $r_m$ . Substituting  $\sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl}$  for the mean and 1 for the scale parameter derives equation 2.3 from equation 2.1:

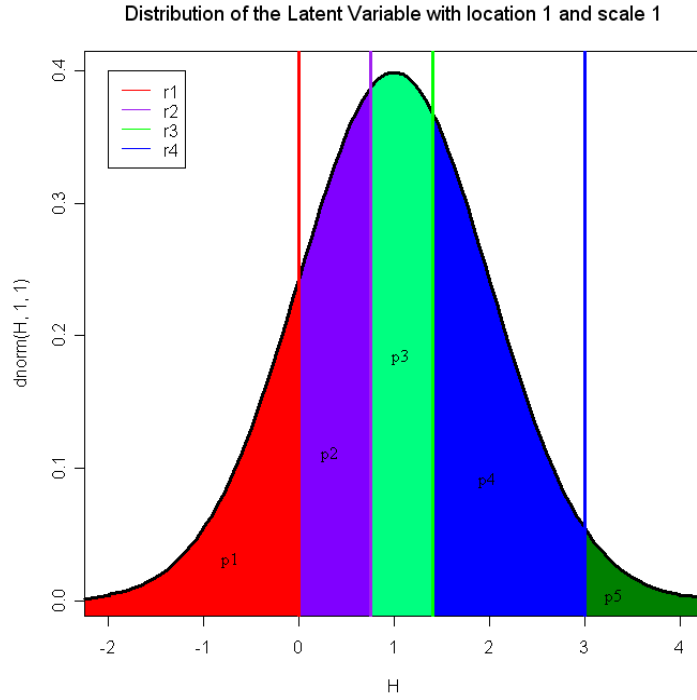


Figure 2.1: Distribution of the Latent Variable. This distribution of the latent variable is divided into five intervals. The vertical lines r1 to r4 represent the cutoff parameters that separate the different possible responses. The areas p1 to p5 correspond to the probability of an observed rating of 0, 1, 2, 3, or 4, respectively.

$$p_{ijlm} = \int_{r_{m-1}}^{r_m} \frac{e^{-\left[ z_{ijl} - \left( \sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl} \right) \right]}}{\left\{ 1 + e^{-\left[ z_{ijl} - \left( \sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl} \right) \right]} \right\}^2} . \quad (2.3)$$

Because the observed responses were mutually exclusive and discrete, the likelihood formed a multinomial distribution with probabilities  $p_{ijlm}$ :

$$L(y | p) \propto \prod_{i=1}^I \prod_{j=1}^J \prod_{l=1}^L \prod_{m=1}^5 p_{ijlm}^{I(y_{ijl}=m)} . \quad (2.4)$$

The indicator variable allowed only one rating by a student on a course for an item.

When student  $i$  in course  $j$  gave rating  $m$  on item  $l$  of the teacher evaluation,  $p_{ijlm}$  appears in the likelihood equation. Johnson did not need to calculate the normalizing constants of

the multinomial distribution to simulate from the posterior distribution, so they do not appear in equation 2.4.

Substituting  $p_{ijlm}$  from equation 2.3 into equation 2.4 yields

$$L(y | Z, r, \beta, u) \propto \prod_{i=1}^I \prod_{j=1}^J \prod_{l=1}^L \int_{r_{y_{ijl-1}}}^{r_{y_{ijl}}} \frac{e^{-(z_{ijl} - \mu_{ijl})}}{[1 + e^{-(z_{ijl} - \mu_{ijl})}]^2} \cdot \quad (2.5)$$

The limits  $r_{y_{ijl-1}}$  and  $r_{y_{ijl}}$  were included because only the  $p_{ijlm}$  that corresponds with the observed rating appears in the likelihood. Note that  $\mu_{ijl}$  is the mean of the latent variable for student  $i$  in course  $j$  on item  $l$ , as modeled above in equation 2.3. Equation 2.5 takes a lot of computation time to integrate the probability density function. The CDF of the logistic density function (F) decreased this time by removing the need to integrate the density function. The integral from  $r_{y_{ijl-1}}$  to  $r_{y_{ijl}}$  equals the difference of the CDFs. The likelihood equation, expressed in terms of the effects and cutoff values with the CDF is

$$L(y | r, \beta, u) \propto \prod_{i=1}^I \prod_{j=1}^J \prod_{l=1}^L \left\{ F \left[ r_{y_{ijl}} - \left( \sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl} \right) \right] - F \left[ r_{y_{ijl-1}} - \left( \sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl} \right) \right] \right\}. \quad (2.6)$$

This formulation of the likelihood excludes the latent variable ( $z_{ijkl}$ ); however, the estimated mean of the latent variable,  $\sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl}$ , does affect the likelihood.

The observed ratings only affected the likelihood through the cutoff parameters.

Then, to make interpretation of the parameters easier, Johnson used  $\theta_m$  to represent the probability that  $y$  ranked in a category below  $m$ . This formulation yielded

$$\theta_m = F \left[ r_m - \left( \sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl} \right) \right]. \quad (2.7)$$

The odds ratio then became  $\frac{\theta_m}{1 - \theta_m} = e^{r_m - (\sum_{k=1}^K x_k \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w_{ijl} u_{ijl})}$ . Johnson reported his results in terms of the proportional odds ratio calculated by

$$\left( \frac{\theta_{m1}}{1 - \theta_{m1}} \right) / \left( \frac{\theta_{m2}}{1 - \theta_{m2}} \right) = e^{r_m - u_1 - (r_m - u_2)}, \quad (2.8)$$

where  $\theta_{m1}$  is the probability that an individual with characteristic 1 rates category  $m$  or below,  $\theta_{m2}$  is the probability that an individual with characteristic 2 rates category  $m$  or below,  $r_m$  is the cutoff parameter for category  $m$ ,  $\mu_1$  is the mean of the latent variable for an individual with characteristic 1, and  $\mu_2$  is the mean of the latent variable for an individual with characteristic 2.

Because  $u_1 - u_2 = \beta_1 - \beta_2$  for individuals who only differ in one characteristic, this equation reduces to  $e^{\beta_1 - \beta_2}$ . Note that the subscript  $m$  does not appear in the reduced form of this equation. This model implies that the proportional odds ratio remains constant for every category of the response variable. For example, the probability that a student who received an A in course  $j$  rates item  $l$  higher than a student in the same course who received a B is reduced to  $e^{(\beta_A - \beta_B)}$ . This probability applies across all rating values.

### 2.3 Possible Prior Distributions for a Bayesian Ordinal Regression Model

Johnson (2002, 2003) did not explicitly state his priors on the course effects, grade effects, or cutoff values in the DUET study; he probably used improper priors based on the simplified computations. However, Johnson did discuss potential prior distributions in *Ordinal Data Modeling* (1999). He first discussed the Jeffery's priors for the effects and the cutoff values. These improper priors do not have a distribution.

However, the cutoff values were constrained so that  $r_{m-1} < r_m$ . The effects ( $\beta$  and  $u$ )

allowed values from negative infinity to positive infinity. The cutoff values,  $r$ , were assumed to be independent of the effects.

Johnson also discussed informative priors. He did not recommend direct specification of the joint prior on  $(\beta, u, r)$  because of the indirect effect of these parameters on the probability of each response category, which demonstrates the difficulty of describing the effects of these parameters individually. He concluded that a combination of the parameters developed by Bedrick, Christensen, and Johnson (1996) is easier to specify because the combination has a more measurable meaning. They specified the cumulative prior probabilities  $\theta$  with a beta density function, and then used the beta density function to bound the probability between 0 and 1 and to allow more flexibility. The parameters of the beta density function also had unique meanings. The researchers decided how many data points their prior was worth. If they did not have much knowledge about the subject, then their prior estimate had less weight compared to the data, while if they were confident about the prior distribution, they increased the weight of the prior. The prior distribution was weighted  $\alpha + \kappa$  data points. The  $\alpha$  parameter corresponded to the number of  $\alpha + \kappa$  data points in category  $m$  or below. Thus,  $\kappa$  represented the number of  $\alpha + \kappa$  data points above category  $m$ . They created one prior cumulative distribution for each of the  $m$  categories, given by

$$\pi(\theta) \propto \prod_{m=1}^M \theta_m^{\alpha_m - 1} (1 - \theta)^{\kappa_m - 1} . \quad (2.9)$$

Using change of variable techniques, they reparameterized the prior into terms of the cutoff values and effects by substituting equation 2.7 for  $\theta_m$  in equation 2.9 and multiplying by the Jacobian. The combined prior distribution resulted in

$$\pi(\beta, u, r) \propto \prod_{m=1}^5 \left( F \left[ r_m - \left( \sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl} \right) \right]^{\alpha_m - 1} \right. \\ \left. \left\{ 1 - F \left[ r_m - \left( \sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl} \right) \right] \right\}^{\kappa_m - 1} f \left[ r_m - \left( \sum_{k=1}^K x' \beta_{kl} + \sum_{i=1}^I \sum_{j=1}^J w' u_{ijl} \right) \right] \right). \quad (2.10)$$

Because this prior results in a complicated posterior distribution with no closed form, the uninformative priors are more commonly used.

#### 2.4 MCMC and Cowles' Hybrid Metropolis-Hastings Algorithm

The posterior distribution is not a known distribution, so MCMC methods must be used to simulate from this distribution. These chains were first used in physics to generate simulations from integrals that could not be solved analytically (Metropolis et al. 1953). W. K. Hastings (1970) first used this technique to create a random draw from a posterior distribution. He simulated candidate values for the posterior using a symmetric function. Then he compared the ratio of the likelihood evaluated at the previous iteration to the candidate values. If the likelihood ratio was greater than 1, the candidate values were accepted in this iteration; otherwise, the candidate values were accepted with the probability of the likelihood ratio. Hastings showed that this process formed a stationary state equivalent to the posterior distribution.

Hastings (1970) also applied this method to multidimensional models. In this instance, he considered one parameter at a time and treated all other parameters as constants. Any parameters that could be separated from this parameter as a multiplicative factor canceled in the numerator and denominator of the likelihood ratio. The resulting equation is called a complete conditional. He simulated draws from the complete conditionals sequentially. The later complete conditionals used the results of the previous complete conditionals. Once a new value was generated for each complete



conditional, he started the next iteration. Although the posterior distribution could not be analytically solved, he showed that this algorithm converged to the steady state found in the posterior distribution. This process was later named the Metropolis-Hastings algorithm.

In ordinal regression, the complete conditionals have closed-form solutions when using the improper priors. Johnson showed that the complete conditionals of the ordinal regression for the latent variables had independent, truncated logistic distributions (1999). Furthermore, he explained that the complete conditional for each cutoff parameter was uniformly distributed between the two adjoining cutoff parameters. Johnson noted the small size of these complete conditionals when there are a large number of observations in adjacent categories. This small size inhibited the movement of the cutoff parameters. He suggested a hybrid of the Metropolis-Hastings algorithm developed by Cowles to calculate the joint conditional distribution of the cutoff parameters and the latent variables (1996). Cowles used the following equation to calculate the joint conditional distribution:

$$\pi(Z, r | y, \beta) = \pi(Z | y, r, \beta)\pi(r | y, \beta). \quad (2.11)$$

The first part of the right hand side of the equation,  $\pi(Z | y, r, \beta)$ , represents the complete conditional of the latent variable previously discussed. The latent variables were simulated from this distribution and accepted with the same probability as the cutoff parameters. The second part of the right hand side of the equation,  $\pi(r | y, \beta)$ , was the likelihood with the latent variable integrated out, as found in equation 2.6. The cutoff parameters moved more and converged sooner when the latent variables and cutoff parameters were updated together. Cowles decreased the logistic model's "burn time," or

the number of iterations until the posterior distribution was reached, from over 1000 iterations to 400 iterations.

The complete conditional of the effect parameters was not in closed form with the logistic distribution:

$$\pi(\beta | Z, r, y) \propto \prod_{n=1}^N f(Z_n - x'_n \beta). \quad (2.12)$$

A simple Metropolis-Hastings step was used in this case. The effects were estimated by alternating between Cowles' algorithm, used for the cutoff parameters and the latent variable, and the simple Metropolis-Hastings step, used for the effect parameters. The resulting distribution was the posterior distribution of the parameters.

The Jeffery's priors must be used because an informative prior breaks down the hybrid Metropolis-Hastings algorithm. The key issue of the inability of Cowles' algorithm to accommodate improper priors is that the latent variable cannot be integrated out of the prior. The latent variable can be integrated out of the likelihood (see equation 2.6); however, the latent variable's informative prior distribution cannot be integrated out. Without the ability to integrate out the latent variable from the informed prior, the second part of Cowles' hybrid equation ( $\pi(r | y, \beta)$ , see equation 2.11) cannot be computed. Instead of using Cowles' algorithm, the cutoff parameters and latent variables would each need to have their own Metropolis-Hastings algorithm. Cowles (1996) showed that these Metropolis-Hastings steps would slow down computation time and the convergence rate. This project will assume the improper prior to use the hybrid Metropolis-Hastings algorithm.

### 3. METHODOLOGY

#### 3.1 Formulation of the Bayesian Ordinal Regression Model

The data from the ICAP survey was analyzed using ordinal regression techniques. The ICAP model resembles Johnson's grade model, but does not have a random effect for each individual, simplifying the model. The ICAP model includes more fixed effects than the grade model. The analysis model examines the affect that gender, disability, mental retardation level, ethnicity, and marital status have on maladaptive behavior. Because the ICAP provides 13 categories for disability, for subjects with multiple disabilities the primary disability is used in the analysis. Utah's Department for Disability Services provided a large sample of 7,772 people for this dataset. The analysis only uses 7,687 of these applicants because 85 subjects have missing data in the relevant explanatory variables.

The design matrix consists of 7,687 rows and 28 columns (one for the intercept, one for gender, twelve for disability, five for mental retardation level, five for ethnicity, and four for levels of marital status). The ordinal regression model uses the effects model rather than the cell means model by including an intercept. In this model,  $\beta_0$  represents a white male with a primary disability of mental retardation, a mild mental retardation level, and a marital status of never married. This parameterization minimizes the error of the intercept because it contains the maximum sample size. Tables 3.1–3.5 contain the observed frequency distributions of the explanatory variables.

Table 3.1: Frequency of Categories of Gender

Gender	Frequency
Male	4357
Female	3330
Total	7687

Table 3.2: Frequency of Categories of Disability

Disability	Frequency
None	17
Autism	508
Blindness	27
Brain damage	515
Cerebral palsy	504
Chemical dependency	8
Deafness	20
Epilepsy	138
Mental retardation	4788
Physical health requiring a nurse	58
Mental illness	97
Situational mental health problems	43
Other	964
Total	7687

Table 3.3: Frequency of Categories of Mental Ability

Mental Level	Frequency
Not mentally retarded	628
Mild	2626
Moderate	1243
Severe	775
Profound	871
Unknown	1544
Total	7687

Table 3.4: Frequency of Categories of Ethnicity

Ethnicity	Frequency
White	7114
Black	74
Oriental	81
Native American	131
Hispanic	228
Other	59
Total	7687

Table 3.5: Frequency of Categories of Marital Status

Marital Status	Frequency
Never married	7245
Married	198
Separated	32
Divorced	191
Widow or widower	21
Total	7687

The parameters of interest are denoted by  $\beta$ . For example,  $\beta_1$  represents the odds ratio of a female having a higher score than a male. The response variables come from the SIB-R section of the ICAP. The SIB-R uses a Likert scale from 0 to 4 for severity and 0 to 5 for frequency. Because the cutoff value to distinguish 0 and 1 is fixed at 0, only three cutoff parameters are estimated for the five categories in the severity model. The frequency model estimates four cutoff parameters for the six categories. Finally, each model has a vector of latent variables with the same length as the response (7,687).

### 3.2 Simulation from the Complete Conditional Distributions

Improper priors are assumed, so the hybrid Metropolis-Hastings algorithm developed by Cowles can be used. Cowles' algorithm generates values for the effect parameters from the joint distribution of the latent variables and the cutoff parameters. This algorithm will speed up convergence and allow the cutoff parameters to move more freely (Cowles 1996). A more thorough description of this algorithm is included below.

First, the parameters are initialized to starting values. All the effect parameters and latent variables were initialized to 1. For this project, the cutoff parameters were initialized to 1, 2, and 3 respectively. Following initialization, the Metropolis-Hastings algorithm generates candidate values from a normal distribution with the mean of the old parameter value and standard deviation of the candidate sigma. Each cutoff parameter,  $r_m$ ,

is truncated from  $r_{m-1}$  to  $r_{m+1}$ . These new cutoff parameters are accepted with the following probability:

$$P = \text{MIN} \left( \prod_{n=1}^N \frac{F(c_{y_n} - x'_n \beta) - F(c_{y_{n-1}} - x'_n \beta)}{F(r_{y_n} - x'_n \beta) - F(r_{y_{n-1}} - x'_n \beta)} \prod_{m=2}^M \frac{F[(r_{m+1} - r_m) / \sigma_c] - F[(c_{m-1} - r_m) / \sigma_c]}{F[(c_{m+1} - c_m) / \sigma_c] - F[(r_{m-1} - c_m) / \sigma_c]}, 1 \right), \quad (3.1)$$

where  $F$  is the cumulative density function of the logistic distribution,  $\sigma_c$  is the standard deviation of the distribution of the candidate value,  $N$  is the number of observations in the data,  $M$  is the number of possible ratings,  $y_n$  is the observed rating for the  $n^{\text{th}}$  observation,  $x'_n$  is the row of indicator variables corresponding to the  $n^{\text{th}}$  observation,  $\beta$  is the column of fixed effect parameters from the previous iteration,  $c_{y_n}$  is the candidate value for the upper cutoff parameter for observation  $y_n$ ,  $c_{y_{n-1}}$  is the candidate value for the lower cutoff parameter for observation  $y_n$ ,  $r_{y_n}$  is the value from the previous iteration for the upper cutoff parameter for observation  $y_n$ ,  $r_{y_{n-1}}$  is the value from the previous iteration for the lower cutoff parameter for observation  $y_n$ ,  $c_m$  is the candidate value for the  $m^{\text{th}}$  cutoff parameter, and  $r_m$  is the value from the previous iteration for the  $m^{\text{th}}$  cutoff parameter. The first half of the equation comes from the complete conditional of the latent variable's joint distribution. The second half comes from the distribution of cutoff parameters given the effect parameters shown in equation 2.11.

If the candidate values for the cutoff parameters are accepted, new values for the latent variables need to be generated. However, if the values for the cutoff parameters are not accepted, the latent variables do not change for the next iteration. Values for the latent variable are drawn from the complete conditional

$$\pi(z_n | r, \beta, y) = f(x'_n \beta) I(r_{y_{n-1}}, r_{y_n}), \quad (3.2)$$

where  $z_n$  is the latent variable for the  $n^{\text{th}}$  observation,  $f$  is the pdf of the logistic distribution, and  $I$  is the indicator function used to truncate the logistic distribution.

Finally, the fixed effects are updated with a Metropolis-Hastings step. The candidate values are drawn from a multivariate normal distribution with the mean of the fixed effects from the last iteration and variance of  $(X'X)^{-1}$ . The acceptance probability for the candidate values of the fixed effects is the ratio of two likelihoods,

$$P = \text{MIN} \left( \prod_{n=1}^N \frac{f(z_n - x'_n C)}{f(z_n - x'_n \beta)}, 1 \right), \quad (3.3)$$

where  $C$  is the vector of the candidate values for the fixed effects.

The expected length to convergence is 500 iterations (Cowles 1996). A burn of 2,000 iterations was used in this project. Ten thousand draws were generated from the complete conditional distribution of each parameter. The mixing plots were checked for convergence and for coverage of the posterior distribution. The expected effects of the explanatory variables were evaluated by creating credible intervals for the parameters of interest. These intervals give a range of reasonable values for the odds ratios of the effects on maladaptive behavior scores. For example, if the credibility interval for the difference between the single (never married) coefficient and the married coefficient does not include 1, then the odds ratio is considered statistically significant. With these intervals, conclusions were drawn about the effect of different demographics on the frequency and severity of maladaptive behavior among disabled individuals.

## 4. RESULTS

### 4.1 Validity of the Frequency Models

Mixing plots are used to examine convergence of MCMC samples. Selected representative mixing plots are included in this chapter. The “Hurtful to Self” problem behavior is used to illustrate the analysis method used for all the problem behaviors. The raw data is plotted to allow comparison to the statistical models. Every model contains significant fixed effects even with the high variation among individual disabled children. The significant parameters are generally consistent across multiple problem behaviors.

An example of the mixing plots for the “Hurtful to Self” problem behavior of the estimated cutoff parameters and the intercept for the frequency are found in Figures 4.1–4.2. These plots are illustrative of mixing in all the models.

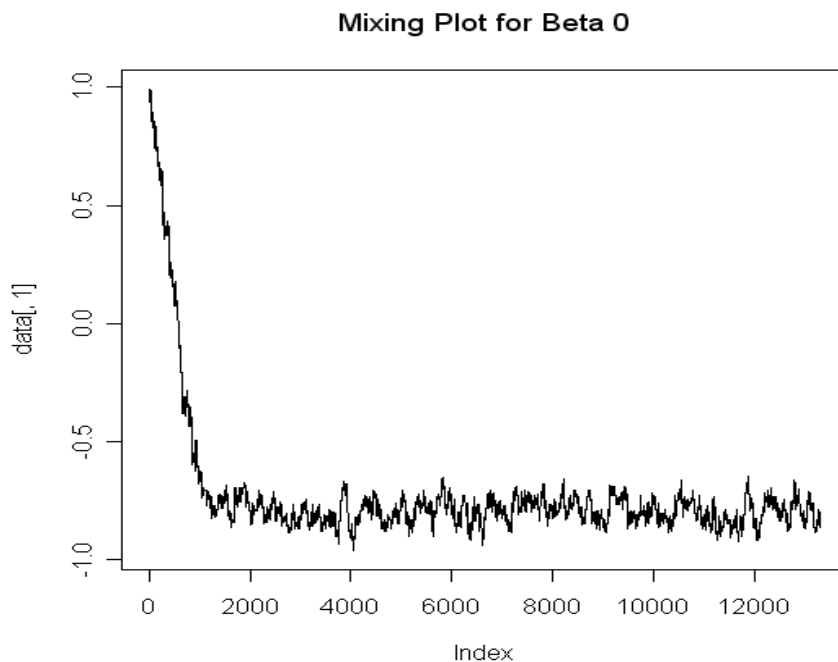


Figure 4.1: Mixing Plot for the Intercept. This plot shows the simulated values of the intercept while the iterations increase. The simulated values appear to have converged to the posterior distribution of the intercept by 2,000 iterations.



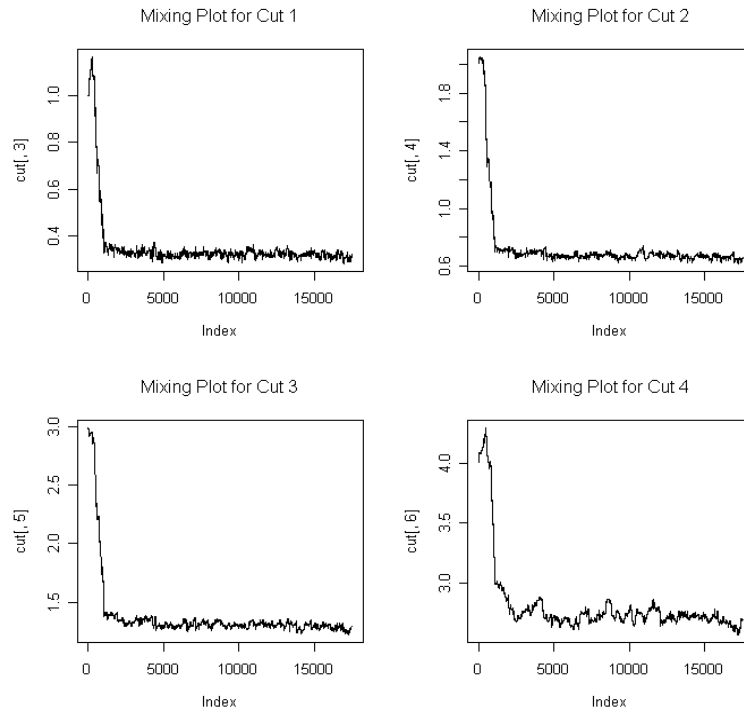


Figure 4.2: Mixing Plots for the Cutoff Parameters. These plots show the simulated values of the cutoff parameters while the iterations increase. The simulated values appear to have converged to the appropriate distribution by 2,000 iterations.

The mixing plot of the samples from the posterior of the gender parameter is shown in Figure 4.3. Gender is not significant in the “Hurtful to Self” model because the credible interval of the odds ratio includes 1 (Figure 4.4). The marginal frequency distributions of the raw data for the two genders show that the counts are virtually identical at each level (Figure 4.5).

The mixing plot for autism is shown in Figure 4.6. The autism parameter is an example of a significant odds ratio. Because this odds ratio is greater than 1, those with autism are at a significantly higher risk for frequent “Hurtful to Self” behavior (Figure 4.7).

An example of lower risk is the cerebral palsy parameter. The mixing plot is shown in Figure 4.8. Figure 4.9 shows the odds ratio centered around .6, so a child with

cerebral palsy has a significantly lower risk of “Hurtful to Self” behavior. A plot of the raw data is found in Figure 4.10, in which autism has a lower frequency of zero ratings and a higher frequency of high ratings. Cerebral palsy has a lower frequency of high ratings and a higher frequency of low ratings.

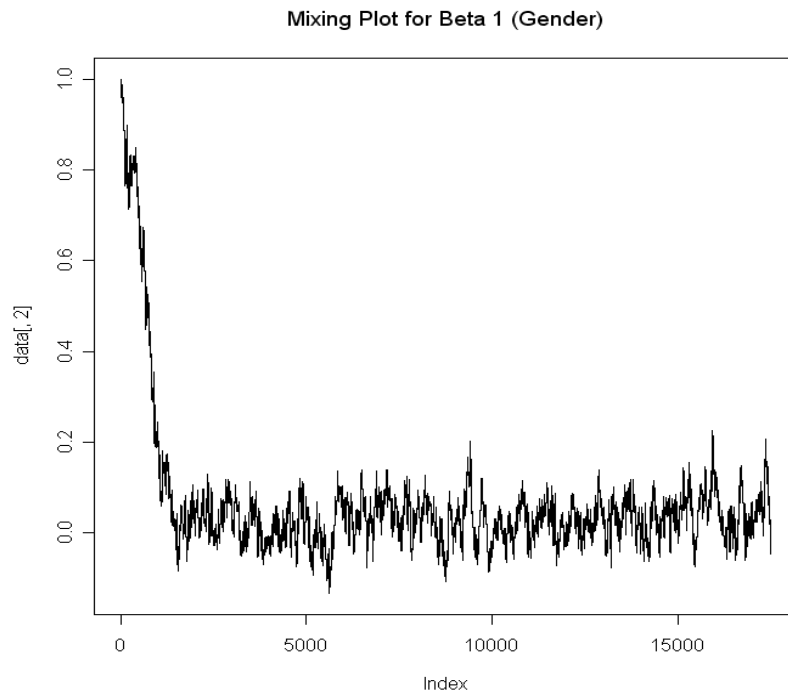


Figure 4.3: Mixing Plot for the Gender Parameter. This plot shows the simulated values of the log difference in the odds ratio while the iterations increase.

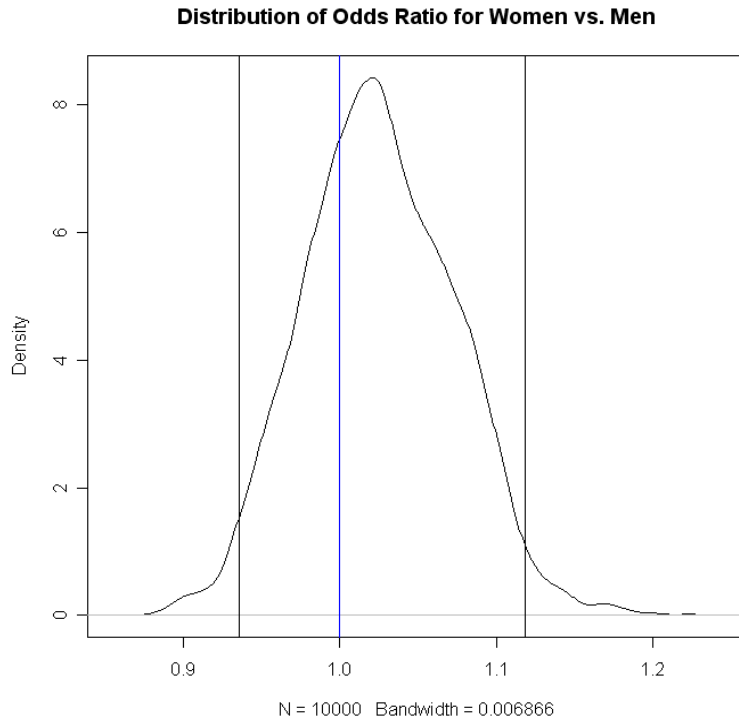


Figure 4.4: Posterior Distribution of the Odds Ratio of Men to Women. This plot shows the posterior distribution of the parameter measuring the effect of gender.

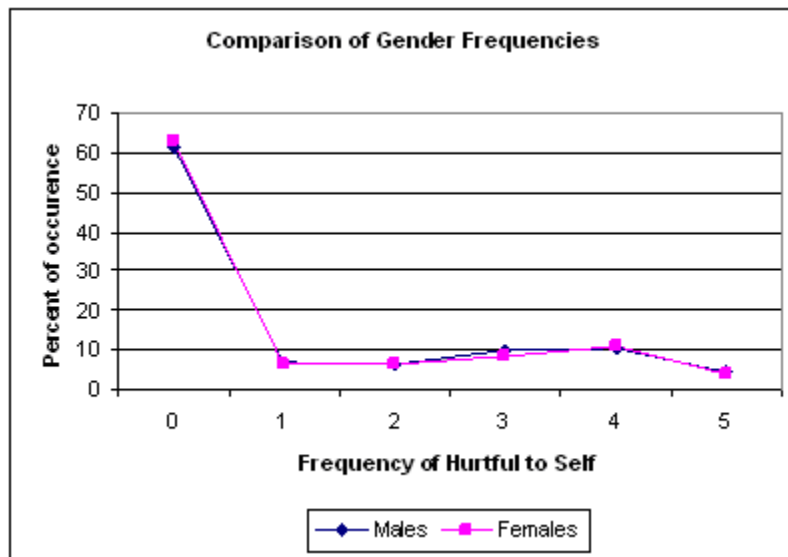


Figure 4.5: Graph of Responses to “Hurtful to Self” by Gender. This line graph shows the frequency of ‘Hurtful to Self’ behavior in the raw data. Males and Females show approximately equal percentages in all levels of the frequency scale

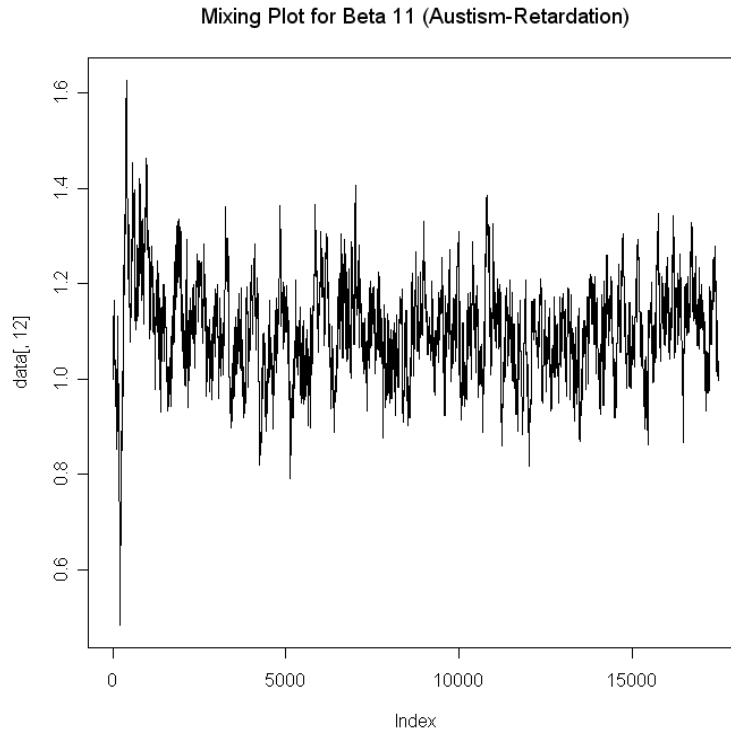


Figure 4.6: Mixing Plot for an Autism Parameter. This plot shows the simulated values of the log difference in the odds ratio while the iterations increase.

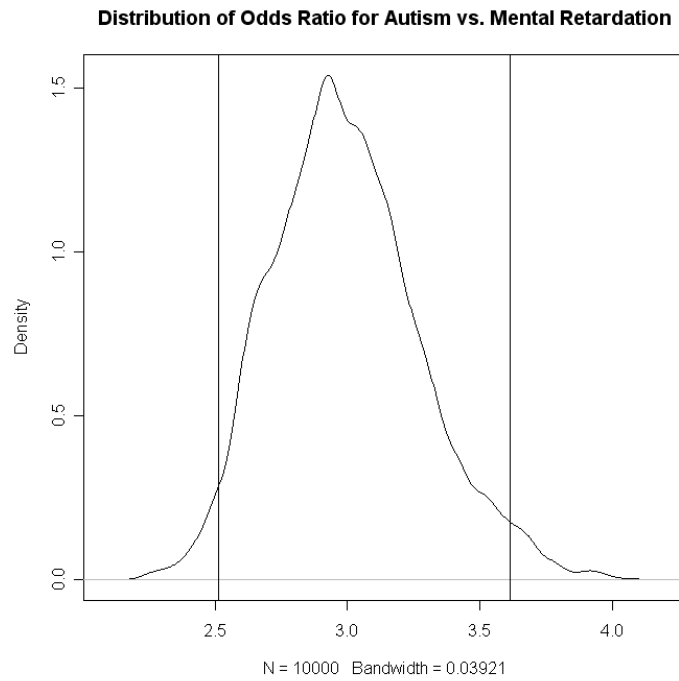


Figure 4.7: Posterior Distribution of the Odds Ratio of Autism to Mental Retardation. This distribution shows that Autistic applicants have a higher risk for frequent “Hurtful to Self” behavior.

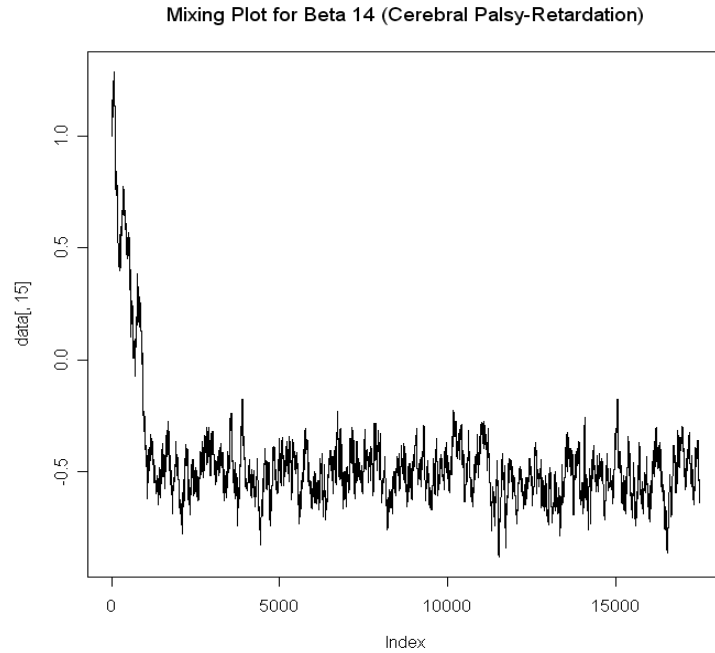


Figure 4.8: Mixing Plot for a Cerebral Palsy Parameter. This plot shows the simulated values of the log difference in the odds ratio while the iterations increase.

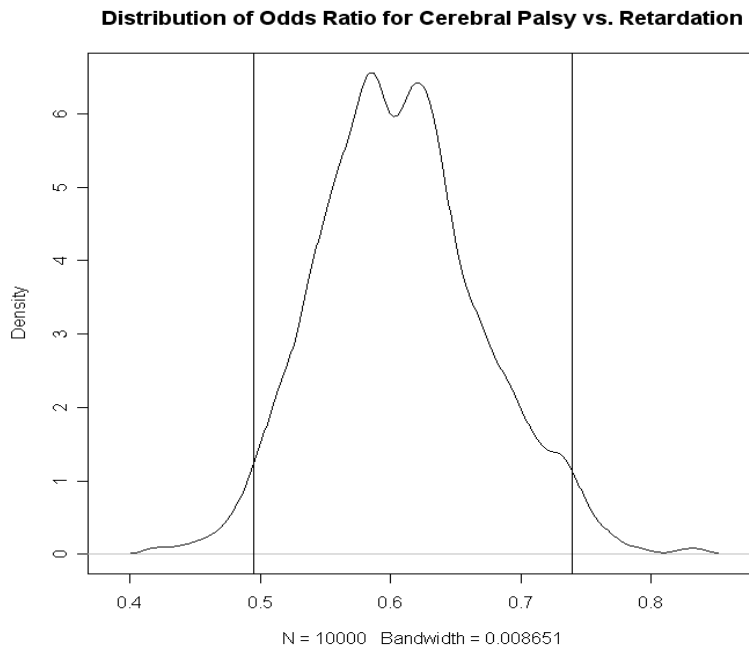


Figure 4.9: Posterior Distribution of the Odds Ratio of Cerebral Palsy to Mental Retardation. This distribution shows that cerebral palsy applicants have a lower risk for frequent “Hurtful to Self” behavior.

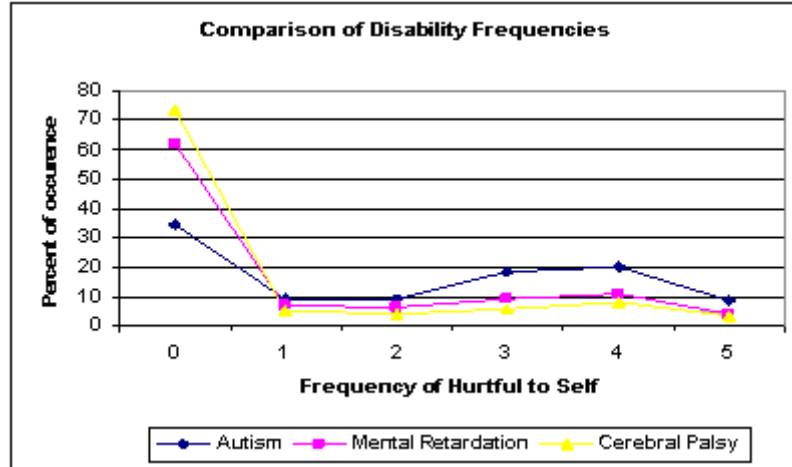


Figure 4.10: Graph of Responses to “Hurtful to Self” by Disability. This line graph shows the frequency of “Hurtful to Self” behavior in the raw data. Autism shows higher percentages in the frequency scale at 3, 4, and 5. Cerebral Palsy shows higher percentages in the 0 rating and lower percentages at the 3, 4, and 5 levels.

## 4.2 Validity of the Severity Models

The severity models are also illustrated using the “Hurtful to Self” problem behavior. Mixing plots for the intercept and the cutoff parameters are shown in Figures 4.11 and 4.12. The severity models have one less cutoff parameter than the frequency models because the scale of the severity ratings ends at 4 instead of 5.

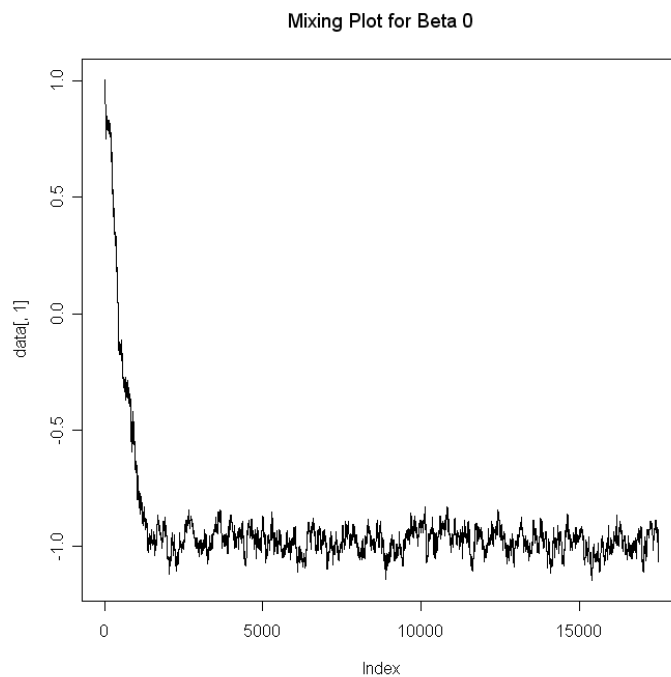


Figure 4.11: Mixing Plot for the Intercept. This plot shows the simulated values of the intercept while the iterations increase. The simulated values appear to have converged to the posterior distribution of the intercept by 2,000 iterations.

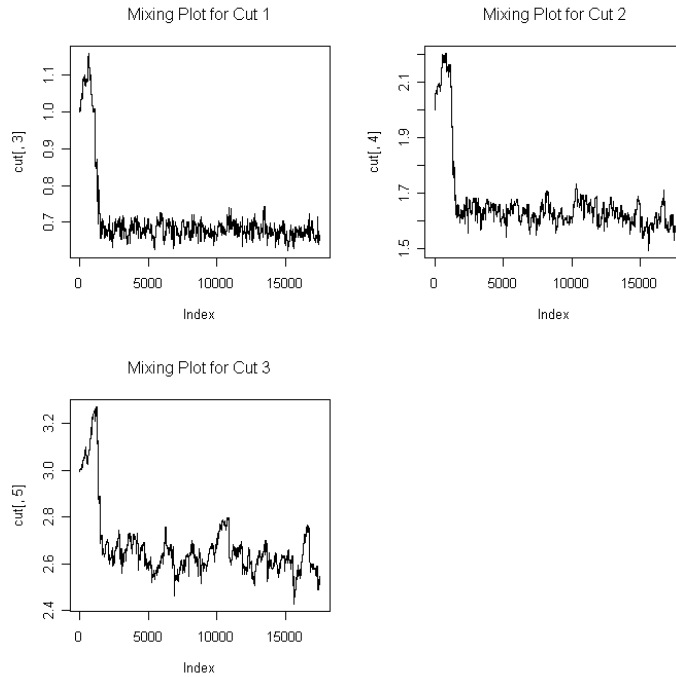


Figure 4.12: Mixing Plots for the Cutoff Parameters. These plots show the simulated values of the cutoff parameters while the iterations increase. The simulated values appear to have converged to the posterior by 2,000 iterations

The mixing plot for the gender parameter is shown in Figure 4.13. Again, gender is not significant in the “Hurtful to Self” model because the posterior distribution of the odds ratio includes 1 (Figure 4.14). The frequency distribution of the raw data is shown in Figure 4.15.



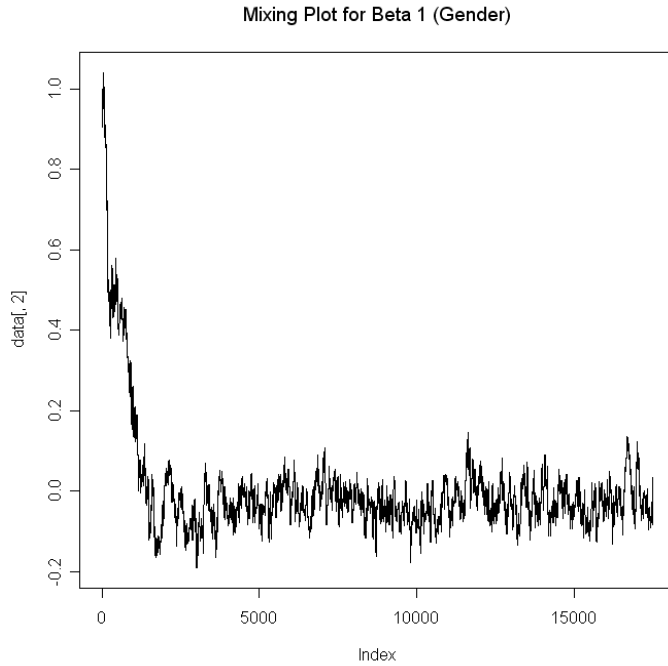


Figure 4.13: Mixing Plot for the Gender Parameter. This plot shows the simulated values of the log difference in the odds ratio while the iterations increase.

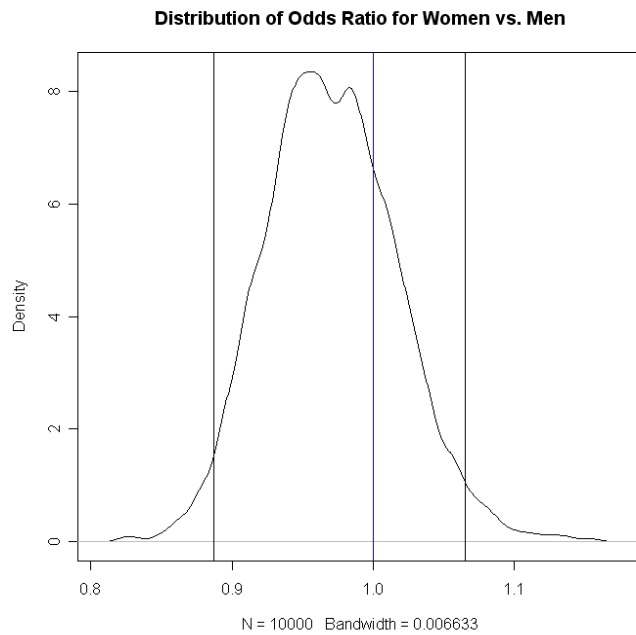


Figure 4.14: Posterior Distribution of the Odds Ratio of Men to Women. This plot shows the posterior distribution of the parameter measuring effect of gender.

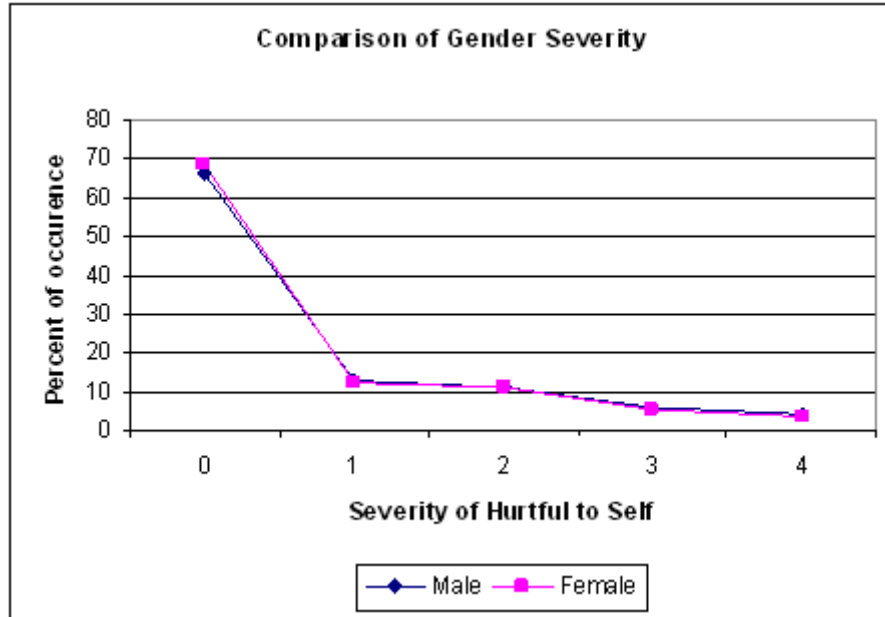


Figure 4.15: Graph of Responses to Hurtful to Self by Gender. This line graph shows the severity of “Hurtful to Self” behavior in the raw data. Males and females show approximately equal percentages in all levels of the severity scale.

The autism parameter shows a higher risk of severe “Hurtful to Self” behavior.

The mixing plot is shown in Figure 4.16. Figure 4.17 shows that the odds ratio for autism to mental retardation is centered around 3.5, and is significantly higher than 1.

The mixing plot for cerebral palsy is shown in Figure 4.18. This parameter’s odds ratio is centered around .5 and also differs significantly from 1. The raw data are shown in Figure 4.20. In this figure, autism has a lower frequency of zero rating and a higher frequency of high ratings. Cerebral palsy has a lower frequency of high ratings and a higher frequency of low ratings.

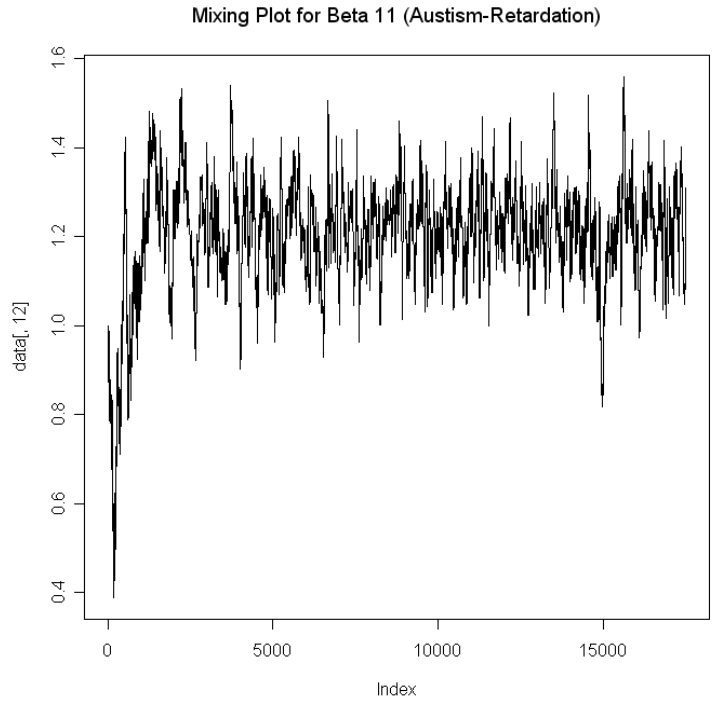


Figure 4.16: Mixing Plot for an Autism Parameter. This plot shows the simulated values of the log difference in the odds ratio while the iterations increase.

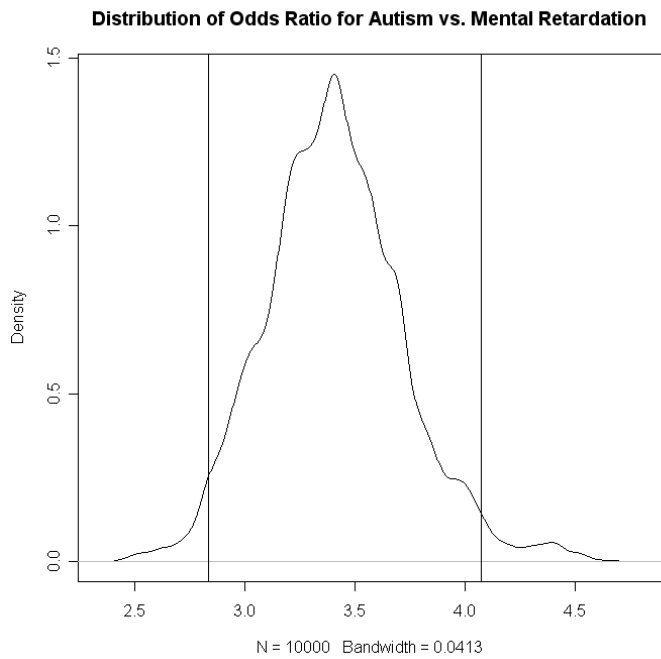


Figure 4.17: Posterior Distribution of the Odds Ratio of Autism to Mental Retardation. This distribution shows that autistic applicants have a higher risk of severe “Hurtful to Self” behavior than those suffering from mental retardation.

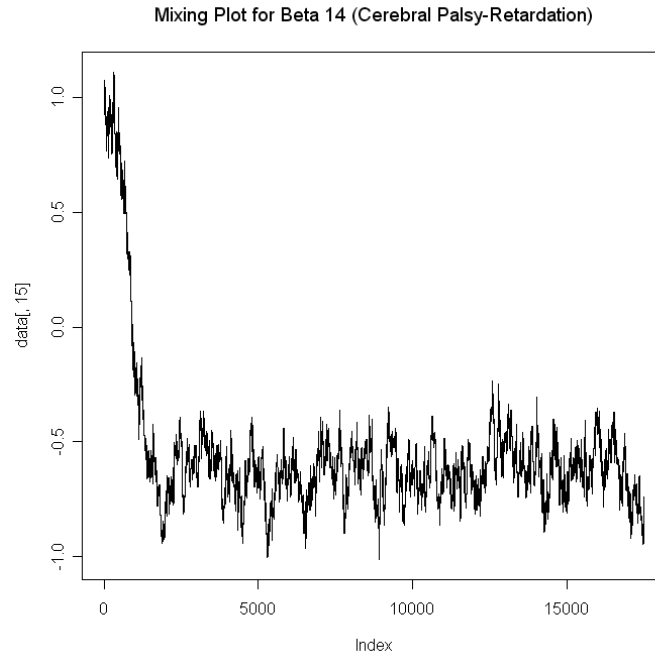


Figure 4.18: Mixing Plot for a Cerebral Palsy Parameter. This plot shows the simulated values of the log difference in the odds ratio while the iterations increase.

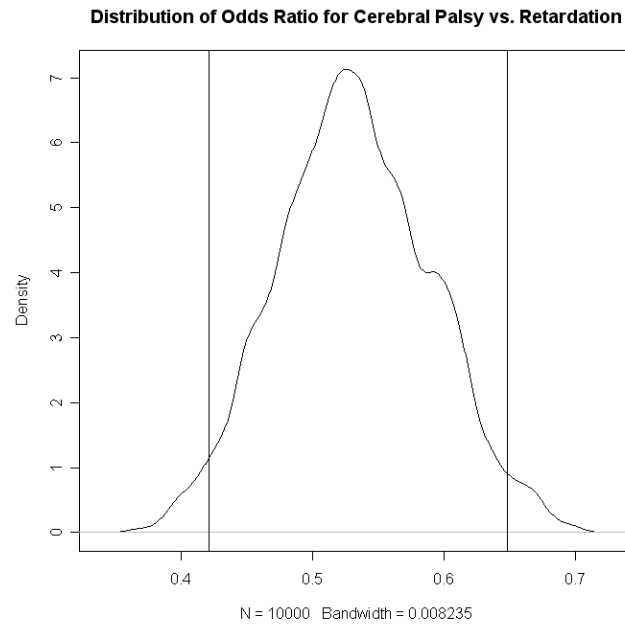


Figure 4.19: Posterior Distribution of the Odds Ratio of Cerebral Palsy to Mental Retardation. This distribution shows that Cerebral Palsy applicants have a lower risk of severe “Hurtful to Self” behavior than those suffering from mental retardation.

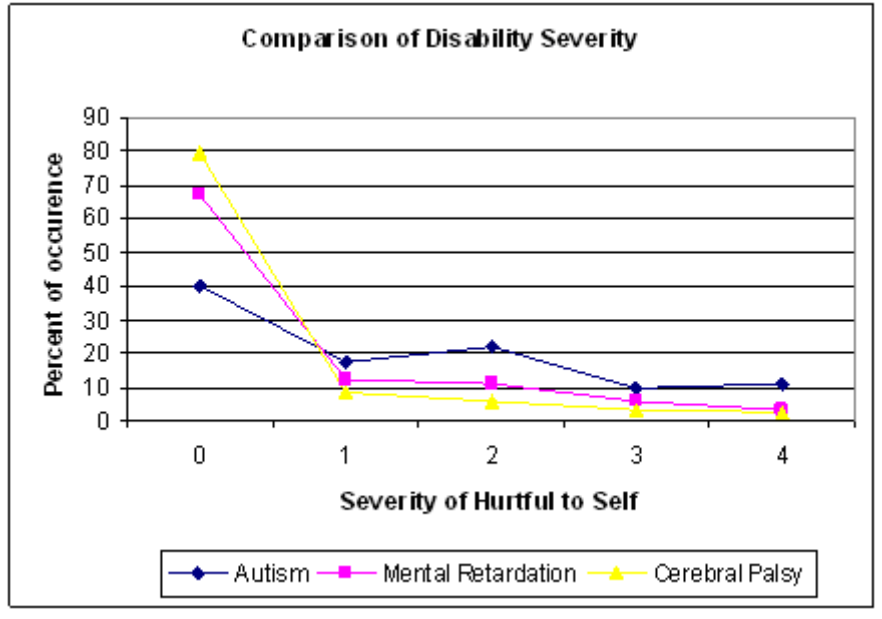


Figure 4.20: Graph of the Responses to Hurtful to Self by Disability. This line graph shows the severity of “Hurtful to Self” behavior in the raw data. Autism shows higher percents in the severity scale at 3, 4, and 5. Cerebral Palsy shows higher percents in the 0 rating and lower percents at the 3, 4, and 5 levels.

### 4.3 Results of the Bayesian Ordinal Regression Models

The results for the rest of the models and parameters are found in Tables 4.1 and 4.2. The lower limits greater than 1 are highlighted in orange to indicate that the odds ratio is significantly greater than 1. The upper limits less than 1 are highlighted in green to indicate that the odds ratio is significantly less than 1. A few interesting trends are discussed below.

First, disability type seems to be an important factor in the risk for frequent and severe problem behaviors. Those with autism exhibited the most risk for frequent and severe maladaptive behavior in all categories except frequency of “Hurtful to Self.” In this category those with profound mental retardation had a higher credible interval than those with autism, but the credible intervals still overlapped. Those with cerebral palsy exhibited significantly lower risk for frequent and severe maladaptive behavior across all problem behaviors. The applicants with physical disabilities requiring care by a nurse had a lower risk of frequent and severe maladaptive scores in all but five of the credible intervals. Blind people had only had one significant credible interval; they displayed more risk for frequent “Unusual and Repetitive Habits” than those with a primary diagnosis of mental retardation. Those with brain injury had a lower risk of severe and frequent “Uncooperative Behavior.” The applicants with a primary diagnosis of chemical dependency only had one significant interval, severity of “Hurtful to Others.” Those with no primary disability had only one significant credible interval, a lower risk of “Unusual or Repetitive Habits” than those with mental retardation. The applicants with epilepsy had a greater risk for severe “Hurtful to Others” and “Destructive to Property” behaviors. Applicants with epilepsy also had a greater risk of frequent “Hurtful to Others.” Lastly,

those with a primary diagnosis of severe mental illness (such as schizophrenia) or with situational mental illness (such as anxiety and depression) had greater risk for frequent and severe problem behaviors. The significant credible intervals in the frequency models for severe mental disability are for “Destructive to Property,” “Unusual or Repetitive Habits,” “Socially Offensive Behavior,” “Inattentive Behavior,” and “Uncooperative Behavior.” All problem behaviors in the severity models had significant credible intervals for severe mental disability. The situational mental illness parameters in the frequency models had significant credible intervals for “Destructive to Property,” “Inattentive Behavior,” and “Uncooperative Behavior.” The severity parameters for situational mental illness had four significant credible intervals: “Hurtful to Self,” “Disruptive Behavior,” “Inattentive Behavior,” and “Uncooperative Behavior.”

The second important characteristic is retardation level. In general, while an applicant’s IQ score lowered, their maladaptive behavior scores increased. Those with severe or profound retardation had significantly greater risk for high frequency ratings in all eight problem behaviors. Interestingly enough, only a few severity ratings are significantly higher than mild retardation. The three significant intervals with a lower limit above 1.05 were “Hurtful to Self,” “Hurtful to Others,” and “Inattentive Behavior.” The applicants with moderate retardation had significant credible intervals in the same three areas. Those with no mental retardation level had significantly lower risk of frequent and severe problem behaviors, the most significant being the severity of “Uncooperative Behavior.”

The third most significant characteristic of the model is marital status. Those applicants who are currently married had the lowest maladaptive behavior scores

(significant credible intervals in frequency and severity of all problem behaviors except “Inattentive Behavior”). Those who are widowed also exhibited lower risk for maladaptive behavior in several areas, including “Destructive to Property,” “Socially Offensive Behavior,” and “Uncooperative Behavior.” They also had a significant credible interval for the frequency of “Unusual or Repetitive Habits.” Those who are married but separated had no significant credible intervals. The divorced applicants also had a lower risk of problem behavior in multiple areas, including the severity of “Hurtful to Others,” and “Destructive to Property,” and the frequency of “Hurtful to Others,” “Destructive to Property,” and “Unusual or Repetitive Habits.” They also had one significantly higher credible interval in the severity of “Inattentive Behavior.” Overall, the applicants with the most severe and frequent problem behaviors were never married at any time.

Gender and Ethnicity also played a role in the maladaptive behavior scales. Females had significantly lower credible intervals for the frequency and severity of almost all problem behaviors. The frequency and severity of “Hurtful to Self” were not significant. Also, the frequency of “Disruptive Behavior” was not significant. Native Americans were at greater risk for severe and frequent “Uncooperative Behavior.”



Table 4.1: Credible Intervals for the Parameters in the Frequency Models

Credible Interval on Frequency of Problem Behavior																	
Demographic	Parameter	Unusual or Socially Offensive Inattentive Uncooperative Behavior															
		Hurtful to Self		Hurtful to Others		Destructive to Property		Disruptive Behavior		Repetitive Habits		Socially Offensive Behavior		Inattentive Behavior		Uncooperative Behavior	
		LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL
Gender	Female	0.9	1.1	0.7	0.8	0.6	0.7	0.9	1.0	0.8	1.0	0.7	0.9	0.8	1.0	0.8	0.9
Ethnicity	Black	0.6	1.6	0.6	1.3	0.4	1.2	0.6	1.4	0.7	1.7	0.4	1.1	0.6	1.6	0.6	1.4
	Asian	0.6	1.5	0.9	1.9	0.9	2.1	0.8	1.6	0.7	1.7	0.6	1.3	0.7	1.7	0.6	1.5
	Native American	0.7	1.5	0.9	1.9	0.9	1.8	0.7	1.2	0.6	1.0	0.6	1.2	0.6	1.2	1.0	1.9
	Other Ethnicity	0.7	1.9	0.4	1.2	0.6	1.6	0.6	1.4	0.9	2.2	0.9	2.2	0.9	2.6	0.7	1.8
	Hispanic	0.7	1.2	0.7	1.3	0.6	1.2	0.6	1.1	0.8	1.3	0.6	1.0	0.7	1.2	0.6	1.1
Marital Status	Married	0.5	0.9	0.5	0.9	0.4	0.7	0.5	0.8	0.5	0.8	0.5	0.9	0.7	1.1	0.4	0.7
	Separated	0.5	1.7	0.3	1.3	0.3	1.8	0.5	1.6	0.7	2.2	0.8	2.5	0.6	1.9	0.4	1.7
	Divorced	0.6	1.2	0.3	0.7	0.5	0.9	0.7	1.1	0.5	0.9	0.8	1.4	0.9	1.6	0.7	1.3
	Widowed	0.1	1.2	0.1	1.3	0.0	0.5	0.3	1.2	0.1	0.8	0.2	0.8	0.4	1.7	0.1	0.9
Primary Disability	Autism	2.5	3.6	3.7	5.2	3.3	4.7	3.1	4.2	3.9	5.6	2.5	3.4	4.7	6.7	4.0	5.7
	Blind	0.7	3.5	0.8	3.4	0.3	1.9	0.9	4.0	1.2	4.5	0.9	3.6	0.5	1.8	0.5	1.8
	Brain	0.8	1.1	0.9	1.2	0.7	1.1	0.8	1.1	0.7	1.0	0.7	1.0	0.7	1.1	0.7	1.0
	Cerebral	0.5	0.7	0.4	0.6	0.3	0.5	0.4	0.6	0.4	0.6	0.3	0.4	0.3	0.5	0.3	0.4
	Chemical	0.3	4.5	0.9	10.9	0.4	4.4	0.5	8.7	0.2	2.3	0.3	4.8	0.1	1.8	0.2	2.3
	Deafness	0.3	2.3	0.6	3.2	0.2	1.9	0.3	1.5	0.3	1.9	0.3	1.7	0.2	1.4	0.3	1.8
	Epilepsy	0.8	1.7	1.3	2.5	0.9	1.8	1.0	1.8	0.6	1.2	0.8	1.4	0.8	1.6	0.9	1.7
	None	0.1	1.1	0.4	2.5	0.3	2.4	0.2	1.4	0.2	1.3	0.2	1.3	0.3	1.8	0.3	1.8
	Physical	0.4	1.3	0.4	1.1	0.1	0.4	0.2	0.8	0.3	0.8	0.2	0.7	0.2	0.6	0.3	0.8
	Mental	0.9	2.0	0.9	2.0	1.2	2.5	0.9	1.9	1.3	2.8	1.2	2.5	1.7	3.2	1.2	2.4
	Situational	0.8	2.6	0.9	2.8	1.1	3.2	1.0	2.8	0.6	2.0	0.7	2.0	1.0	2.9	1.1	3.3
Other Disability	0.9	1.2	0.8	1.1	0.8	1.1	0.7	1.0	0.9	1.3	0.6	0.8	0.7	0.9	0.8	1.1	
Mental Retardation Level	None	0.7	1.1	0.7	1.0	0.7	1.0	0.8	1.1	0.7	1.0	0.8	1.1	0.9	1.3	0.7	1.0
	Moderate	1.0	1.3	1.1	1.5	0.9	1.2	1.0	1.2	1.3	1.6	1.0	1.2	0.9	1.1	1.0	1.3
	Severe	1.6	2.1	1.5	2.1	1.2	1.6	1.1	1.5	1.7	2.2	1.1	1.5	1.1	1.4	1.2	1.6
	Profound	2.8	3.7	1.7	2.3	1.2	1.6	1.1	1.4	2.8	3.7	1.0	1.3	1.5	1.9	1.1	1.5
	Unknown	1.0	1.4	1.3	1.7	0.9	1.2	0.8	1.0	1.0	1.3	0.5	0.6	0.6	0.8	0.8	1.0

Table 4.2: Credible Intervals for the Parameters in the Severity Models

Credible Interval on Severity of Problem Behavior																	
Demographic	Parameter	Unusual or Socially Inattentive or Uncooperative Behavior															
		Hurtful to Self		Hurtful to Others		Destructive to Property		Disruptive Behavior		Repetitive Habits		Socially Offensive Behavior		Inattentive Behavior		Uncooperative Behavior	
		LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL
Gender	Female	0.9	1.1	0.6	0.8	0.5	0.7	0.8	1.0	0.8	0.9	0.7	0.8	0.8	1.0	0.7	0.9
Ethnicity	Black	0.6	1.7	0.5	1.5	0.5	1.4	0.6	1.3	0.4	1.0	0.4	1.2	0.6	1.5	0.5	1.2
	Asian	0.6	1.7	0.7	1.6	0.8	1.9	0.8	1.7	0.6	1.4	0.4	1.1	0.7	1.5	0.6	1.4
	Native American	0.6	1.2	0.8	1.7	0.8	1.6	0.8	1.6	0.6	1.2	0.8	1.5	0.7	1.5	1.1	2.1
	Other Ethnicity	0.6	1.6	0.5	1.6	0.7	1.6	0.6	1.6	0.7	1.7	0.8	1.9	0.7	2.0	0.7	1.7
	Hispanic	0.7	1.2	0.7	1.3	0.6	1.0	0.6	1.1	0.8	1.3	0.6	1.0	0.7	1.2	0.7	1.2
Marital Status	Married	0.5	0.9	0.5	0.9	0.3	0.6	0.5	0.9	0.4	0.7	0.5	0.9	0.8	1.3	0.4	0.7
	Separated	0.6	2.5	0.5	2.3	0.2	1.4	0.7	2.3	0.6	2.4	0.8	2.8	0.8	2.9	0.7	2.3
	Divorced	0.7	1.3	0.4	0.9	0.5	1.0	0.8	1.3	0.6	1.0	0.9	1.6	1.2	1.9	1.0	1.6
	Widowed	0.1	1.2	0.1	1.2	0.0	0.5	0.2	1.1	0.2	1.3	0.1	0.8	0.4	2.5	0.2	0.9
Primary Disability	Autism	2.8	4.1	2.9	4.2	3.0	4.3	2.8	3.9	3.7	5.4	2.3	3.4	3.9	5.6	2.8	4.3
	Blind	0.9	3.7	0.6	2.6	0.3	1.5	1.0	3.8	1.0	4.4	0.6	2.1	0.5	1.9	0.4	1.7
	Brain	0.7	1.1	0.8	1.3	0.8	1.1	0.8	1.1	0.7	1.0	0.7	1.1	0.7	1.1	0.7	1.0
	Cerebral	0.4	0.6	0.4	0.6	0.3	0.5	0.4	0.5	0.3	0.5	0.2	0.3	0.3	0.5	0.2	0.4
	Chemical	0.5	8.4	1.4	18.4	0.3	4.7	0.9	11.5	0.3	5.4	0.4	6.9	0.2	2.7	0.3	5.2
	Deafness	0.3	2.2	0.5	3.0	0.3	2.1	0.3	1.7	0.4	1.9	0.3	1.8	0.2	1.0	0.3	1.9
	Epilepsy	0.8	1.6	1.2	2.3	1.0	1.8	1.0	1.8	0.6	1.1	0.7	1.4	0.8	1.6	0.8	1.5
	None	0.1	1.4	0.4	3.0	0.3	3.3	0.3	1.8	0.0	0.6	0.2	1.8	0.3	2.1	0.3	2.5
	Physical	0.4	1.4	0.3	1.1	0.1	0.5	0.3	1.0	0.3	0.9	0.2	0.7	0.2	0.6	0.3	0.9
	Mental	1.1	2.6	1.2	2.6	1.6	3.2	1.2	2.5	1.5	3.0	1.3	2.7	1.8	3.6	1.4	3.0
	Situational	1.5	4.6	0.8	3.0	0.9	3.4	1.4	4.1	0.6	1.8	0.8	2.6	1.7	5.4	1.2	4.1
Other Disability	0.9	1.2	0.8	1.0	0.7	1.1	0.6	0.9	0.8	1.1	0.6	0.9	0.7	0.9	0.7	0.9	
Mental Retardation Level	None	0.8	1.1	0.6	1.0	0.6	1.0	0.7	1.0	0.7	1.0	0.7	1.1	0.9	1.2	0.6	0.9
	Moderate	1.0	1.4	1.0	1.3	0.9	1.2	0.8	1.0	1.0	1.3	0.9	1.1	0.7	1.0	0.9	1.2
	Severe	1.5	2.1	1.4	2.0	1.1	1.5	1.0	1.3	1.3	1.7	1.0	1.3	0.8	1.1	1.0	1.4
	Profound	2.5	3.2	1.4	1.8	1.0	1.4	0.8	1.1	1.8	2.3	0.8	1.1	1.0	1.3	0.8	1.1
Unknown	0.9	1.2	0.9	1.3	0.7	1.0	0.6	0.8	0.8	1.1	0.4	0.6	0.5	0.7	0.6	0.7	

## 5. CONCLUSIONS

This project demonstrates the viability of the ordinal regression model.

Developing an ordinal regression model for the Maladaptive Behavior Subscales in the ICAP for each problem behavior individually has many advantages. The results of the models make intrinsic sense, and interpretation from the models is readily explained because the relationships are explained in terms of odds ratios. The model can also borrow strength from categories containing more data to remedy data deficiencies, and allows the researcher to focus on one behavior rather than combining scales to form indices where normality might be reasonably be assumed. For example, if a researcher thought the treatment imposed only affected the “Destructive to Property” behavior, then a model could be built for that behavior alone without adding the other behaviors to the model.

The credible intervals also have a more direct application to individual cases. These intervals do not describe the odds ratio on average, but rather the odds ratio for a single individual. Because the regression model is additive, all the characteristics of an individual can be analyzed together. A researcher can easily find the odds ratio for an unmarried, Asian female with autism and moderate retardation level in comparison to a married, white male with cerebral palsy and profound retardation level. This can be accomplished by finding the corresponding coefficients and adding them together. Furthermore, the result would not be just a point estimate, but a distribution of the odds ratio, which displays the shape and spread.

Before any conclusions can be drawn from this model it is important to note the population and quality of the data. The data are not from a random selection of people in

the United States or even in Utah. These results can only apply to the subpopulation of applicants for aid from Utah's Division of Services for Persons with Disabilities.

Furthermore, respondents had a motivation to rate the individual higher on both the frequency scale and the severity scale because the Maladaptive Behavior scores are one factor in determining aid. This factor could introduce non-sampling bias that cannot be predicted or eliminated with the current data. Thus, all the conclusions of the model have an assumption of response bias. Still, some significant effects are definitely indicated by the model, as discussed in Chapter 4.

This model also opens possibilities for improvements. First, the sixteen models could be combined into one model with added parameters for the correlation between each variable. This combination would allow the credible intervals to become credible regions to account for dependency among the response variables. Also, interaction terms could be added into the model to allow greater flexibility in the modeling. An example of an interaction term that could be added is the gender and marital status interaction, which would allow the effect of marital status on the response variables to change according to gender. A Bayesian factor could be calculated to find a way to judge the significance of model terms and to enhance model selection.

A limitation of the current model is that improper priors are needed to have a reasonable computation time. Informative priors could be added to the model, but then Cowles' hybrid of the Metropolis-Hastings algorithm would not function. Computation time would increase because the regular Metropolis-Hastings algorithm has a much slower convergence rate. Still, this model may function like a springboard to help researchers devise additional tools to accommodate their data analyses.

## BIBLIOGRAPHY

- Adams, John (1997), "Student Evaluations: the Rating Game," *Inquiry*, 1, 10–16.
- Albert, James, and Chib, Siddhartha (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.
- Blum, M. L. (1936), "An investigation of the relation existing between student's grades and their ratings of instructors' ability to teach," *Journal of Educational Psychology*, 27, 217–221.
- Bruininks R. H., Hill B. K., and Weatherman R. F. (1986), *Inventory for Client and Agency Planning (ICAP)*, Allen, Tx: DLM Teaching Resources.
- Bruininks R. H., Hill B. K., Woodcock, R. W., and Weatherman R. F. (1996), *Scales of Independent Behavior – Revised Comprehensive Manual*, Itasca, IL: Riverside Publishing Company.
- Centra, J. A. (2003), "Will Teachers Receive Higher Student Evaluations by Giving Higher Grades and Less Course Work," *Research in Higher Education*, 44(5), 495–518.
- Costin, F., Greenough, W. T., and Menges, R. J. (1971), "Student ratings of college teaching: reliability, validity, and usefulness," *Review of Educational Research*, 41(5), 511–535.
- Cowles, M. K. (1996), "Accelerating Monte Carlo Markov Chain Convergence for Cumulative-link Generalized Linear Models," *Statistics and Computing*, 6, 86–98.
- Eckert, R. E. (1950), "Ways of evaluating college teaching," *School and Society*, 71, 65–69.

- Feldman, K. A. (1976), "Grades and college students' evaluation of their courses and teachers," *Research in Higher Education*, 4, 69–111.
- Garverick, C. S., and Carter, H. D. (1968), "Instructor ratings and expected grades," *California Journal of Educational Research*, 13, 218–221.
- Greenwald, A. G., and Gillmore, G. M. (1997), "Grading leniency is a removable contaminant of Student ratings," *American Psychologist*, 52(11), 1209–1217.
- Guthrie, E. R. (1949), "The evaluation of teaching," *Educational Record*, 30, 109–115.
- Harville, David A., and Mee, Robert W. (1984), "A Mixed Model Procedure for Analyzing Ordered Categorical Data," *Biometrics*, 40, 393–408.
- Hastings, W.K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their applications," *Biometrika*, 57, 97–109.
- Howard, G.S., and Maxwell, S.E. (1980), "Correlation between student satisfaction and grades: a case of mistaken causation," *Journal of Educational Psychology*, 72, 810–820.
- Ishwaran, Hemant, and Gatsonis, Constantine (2000), "A general class of hierarchical ordinal regression models with applications to correlated ROC analysis," *The Canadian Journal of Statistics*, 28, 731–750.
- Johnson, Valen E. (1996), "On Bayesian Analysis of Multirater Ordinal Data: An application to Automated Essay Grading," *Journal of the American Statistical Association*, 91, 42–51.
- Johnson, Valen E. (1999), *Ordinal Data Modeling*, New York: Springer-Verlag New York, Inc.

- Johnson, Valen E. (2002), "Teacher Course Evaluations and Student Grades: An Academic Tango," *Chance*, 15(3), 9–16.
- Johnson, Valen E. (2003), *Grade Inflation*, New York: Springer-Verlag New York, Inc.
- Marsh, H.W., and Roche, L.A. (1997), "Making students evaluations of teaching effectiveness effective," *American Psychologist*, 52, 1187–1197.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, 21, 1087–92.
- McCullagh, Peter (1980), "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society Series B*, 42, 109–142.
- McKinney, Kathleen (1997), "What do Student Ratings Mean," *The National Teaching and Learning Forum*, 7, 1–2.
- Olsen, D. R. (1997), "Grade Inflation: Reality or Myth? Student Preparation Level vs Grades at Brigham Young University 1975–1994," *Association for Institutional Research Annual Forum*, Orlando, Florida. (ERIC Document Reproduction Service No. ED 410880)
- Orsmond, Gael I., Seltzer, Marsha Mailick et al. (2003), "Behavior Problems in Adults With Mental Retardation and Maternal Well-Being: Examination of the Direction of Effects," *American Journal on Mental Retardation*, 108(4), 257–271.
- Powell, R.W. (1977), "Grades, learning and student evaluation of instruction," *Research in Higher Education*, 7, 193–205.
- Remmers, H.H. (1928), "The relationship between students' marks and student attitude toward instructors," *School and Society*, 28, 759–760.

Thompson, James R., Hughes, Carolyn, Schalock, Robert L., Silverman, Wayne, Tassé, Marc J., Bryant, Brian, Craig, Ellis M., and Campbell, Edward M. (2002),  
“Integrating Supports in Assessment and Planning,” *Mental Retardation*, 40(5),  
390–405.



## A. COMPUTER CODE

#The following is my R code used to obtain draws from the posterior distribution of the frequency parameters#

#Each of the eight models had the same code except the X matrix used Freq\* where \* was the problem behavior#

#This code reads in the data#

```
library(MASS)
data<-read.csv("F:/dataint.csv")
data$ID<-as.factor(data$ID)
data$GENDER<-as.factor(data$GENDER)
data$RACE<-as.factor(data$RACE)
data$MAR<-as.factor(data$MAR)
data$DIS<-as.factor(data$DIS)
data$MENT<-as.factor(data$MENT)
```

#This code initializes variables and functions#

```
burn<-0
length<-20
#The following lines initialize the cutoff parameter vector#
cut0<-matrix(-9999,(length+burn),1)
cut1<-matrix(0,(length+burn),1)
cut2<-matrix(1,(length+burn),1)
cut3<-matrix(2,(length+burn),1)
cut4<-matrix(3,(length+burn),1)
cut5<-matrix(4,(length+burn),1)
cut6<-matrix(9999,(length+burn),1)
cut<-cbind(cut0,cut1,cut2,cut3,cut4,cut5,cut6)
```

#This code initializes the candidate vector for the next simulation.#

```
candcut<-cut[1,]
```

This code initializes the latent variable#

```
latent<-matrix(1,1,nrow(data))#substitute (length+burn) in for 1 to output latent variables
```

#The following lines initialize the beta vector #

```
beta<-matrix(1,(length+burn),28)
```

#Initializes the candidate vector for the next simulation.#

```
candbeta<-beta[1,]
```

#Initializes the mixing parameter for the cutoff values#

```
mixcutsig<-numeric()
```

```
mixcutsig[1]<-0.015
```

```
mixcutsig[2]<-0.015
```

```
mixcutsig[3]<-0.015
```

```
mixcutsig[4]<-0.015
```

#Initializes the mixing parameter for the beta vector#

```
mixbetasig<-0.45
```

#Initializes the pdf of the logistic function (f in all equations)#

```
logis<-function(x,mean){exp(-(x-mean))/(1+exp(-(x-mean)))^2}
```

#Initializes the cdf of the logistic function (F in all equations)

```
clogis<-function(x,mean){1/(1+exp(-(x-mean)))}
```

#Initializes the part of the acceptance ratio for one cutoff parameter (seen in the second half of equation 12)#

```
cpratio1<-function(r){(clogis((cut[int,(r+2)]-cut[int,(r+1)]))/mixcutsig[r-1,0]-clogis((candcut[r]-cut[int,(r+1)]))/mixcutsig[r-1,0])/(clogis((candcut[(r+2)]-candcut[(r+1)])/mixcutsig[r-1,0]-clogis((cut[int,(r)]-candcut[(r+1)])/mixcutsig[r-1,0]))}
```

#initializes the acceptance ratio for the latent variable and cutoff parameters combined.#

```

ratio<-numeric()
#initializes the acceptance ratio for the beta vector#
bratio<-numeric()

#create design matrix (effects model because intercept included)#
X<-model.matrix(FREQ1~GENDER+RACE+MAR+DIS+MENT,data)

#This starts writing iterations#
#I output 19 simulations 700 times to come up with the burn of 2,000 plus the 10,000 and some
extras#
for (rep in 1:700){
#Starts calculation iterations#
for(int in 2:(length+burn)){

#Updates cutoffs and latent variables#
cut[int,3]<-cut[(int-1),3]
#Makes sure that the while loop is not bypassed#
candcut[3]<-cut[int,4]
#Simulates a candidate cutoff value from a truncated normal distribution with the upper limit being
the cutoff parameter one higher and the lower limit being the cutoff parameter one lower.#
while(round(abs(candcut[3]-(cut[(int-1),4]+cut[int,2])/2),6)>=round((cut[(int-1),4]-cut[int,2])/2,6)){
candcut[3]<-rnorm(1,cut[(int-1),3],mixcutsig[1])
}

cut[int,4]<-cut[(int-1),4]
candcut[4]<-cut[int,5]
while(round(abs(candcut[4]-(cut[(int-1),5]+cut[int,3])/2),6)>=round((cut[(int-1),5]-cut[int,3])/2,6)){
candcut[4]<-rnorm(1,cut[(int-1),4],mixcutsig[2])
}

cut[int,5]<-cut[(int-1),5]
candcut[5]<-cut[int,4]
while(round(candcut[5],6)<=round(cut[int,4],6)){
candcut[5]<-rnorm(1,cut[(int-1),5],mixcutsig[3])
}

cut[int,6]<-cut[(int-1),6]
candcut[6]<-cut[int,5]
while(round(candcut[6],6)<=round(cut[int,5],6)){
candcut[6]<-rnorm(1,cut[(int-1),6],mixcutsig[4])
}

#Calculate acceptance ratio#
#Calculates the second half of the acceptance ratio found in equation 12 using the function
defined above called cpratio1#
#Calculates the first half of the acceptance ratio found in equation 12#
ratio[int]<-cpratio1(2)*cpratio1(3)*cpratio1(4)*cpratio1(5)
for (i in 1:nrow(data)){
if(data$FREQ2[i]>0){
ratio[int]<-ratio[int]*(clogis(candcut[(data$FREQ2[i]+2)],X[i,]%%beta[(int-1),])-
clogis(candcut[(data$FREQ2[i]+1)],X[i,]%%beta[(int-
1),]))/(clogis(cut[int,(data$FREQ2[i]+2)],X[i,]%%beta[(int-1),])-
clogis(cut[int,(data$FREQ2[i]+1)],X[i,]%%beta[(int-1),]))
}
}
}

```

```

#Replaces the new candidate values for the cutoff parameters with probability of the acceptance
ratio#
test<-runif(1,0,1)
#Generate from a uniform [0,1] distribution to accept with the correct probability#
if(test<ratio[int]) {
  cut[int,]<-candcut
#Replaces the new latent variables with the candidates simulated from the truncated logistic
distribution#
for (i in 1:nrow(data)){
  latent[1,i]<-rlogis(1,t(X[i,])%*%t(t(beta[(int-1,])),1) #replace latent[int,i] for latent[1,i] to
graph latenat variables#
  while(abs(latent[1,i]-
(cut[int,(data$FREQ2[i]+2)]+cut[int,(data$FREQ2[i]+1)])/2)>=(cut[int,(data$FREQ2[i]+2)]-
cut[int,(data$FREQ2[i]+1)])/2){
    latent[1,i]<-rlogis(1,t(X[i,])%*%t(t(beta[(int-1,])),1)
  }
}
}

#This code draws candidate fixed effect parameters#
#Replaces the new betas if the candidate betas are not accepted#
beta[int,]<-beta[(int-1,)]
#Generates new candidate betas from a multivariate normal distribution#
candbeta<-mvrnorm(n=1,beta[(int-1,)],mixbetasig*solve(t(X)%*%X))

#Uses equation 14 to calculate acceptance ratio#
bratio[int]<-1
for (i in 1:nrow(data)){
  bratio[int]<-
bratio[int]*logis(latent[1,i],t(X[i,])%*%t(t(candbeta)))/logis(latent[1,i],t(X[i,])%*%t(t(beta[int,])))
}
#Replaces the new betas with candidate betas with probability of the acceptance ratio#
test<-runif(1,0,1)
if(test<bratio[int]) {
  beta[int,]<-candbeta
}
}

#Outputs the next 19 simulations for the parameters#
write(t(beta[2:(dim(beta)[1]),]),file="C:/Documents and Settings/Paul
Johnson/Desktop/freqbeta1.txt",ncol=28,append=TRUE)
write(t(cut[2:(dim(beta)[1]),]),file="C:/Documents and Settings/Paul
Johnson/Desktop/freqcut1.txt",ncol=7,append=TRUE)
#Stores the last simulation to generate new simulations#
beta[1,]<-beta[int,]
cut[1,]<-cut[int,]
}
#The following is my R code used to obtain draws from the posterior distribution of the severity
parameters#
#Each of the eight models had the same code except the X matrix used Freq* where * was the
problem behavior#

#This code reads in the data#
library(MASS)
data<-read.csv("F:/dataint.csv")
data$ID<-as.factor(data$ID)

```

```

data$GENDER<-as.factor(data$GENDER)
data$RACE<-as.factor(data$RACE)
data$MAR<-as.factor(data$MAR)
data$DIS<-as.factor(data$DIS)
data$MENT<-as.factor(data$MENT)

#This code initializes variables and functions#
burn<-0
length<-20
#The following lines initialize the cutoff parameter vector#
cut0<-matrix(-9999,(length+burn),1)
cut1<-matrix(0,(length+burn),1)
cut2<-matrix(1,(length+burn),1)
cut3<-matrix(2,(length+burn),1)
cut4<-matrix(3,(length+burn),1)
cut5<-matrix(9999,(length+burn),1)
cut<-cbind(cut0,cut1,cut2,cut3,cut4,cut5)
#This code initializes the candidate vector for the next simulation.#
candcut<-cut[1,]
#This code initializes the latent variable#
latent<-matrix(1,1,nrow(data))
#The following lines initialize the beta vector #
beta<-matrix(1,(length+burn),28)
#Initializes the candidate vector for the next simulation.#
candbeta<-beta[1,]
#Initializes the mixing parameter for the cutoff values#
mixcutsig<-numeric()
mixcutsig[1]<-.015
mixcutsig[2]<-.015
mixcutsig[3]<-.015
#Initializes the mixing parameter for the beta vector#
mixbetasig<-.45
#Initializes the pdf of the logistic function (f in all equations)#
logis<-function(x,mean){exp(-(x-mean))/(1+exp(-(x-mean)))^2}
#Initializes the cdf of the logistic function (F in all equations)
clogis<-function(x,mean){1/(1+exp(-(x-mean)))}
#Initializes the part of the acceptance ratio for one cutoff parameter (seen in the second half of
equation 12)#
cpratio1<-function(r){(clogis((cut[int,(r+2)]-cut[int,(r+1)])/mixcutsig[r-1],0)-clogis((candcut[r]-
cut[int,(r+1)])/mixcutsig[r-1],0))/(clogis((candcut[(r+2)]-candcut[(r+1)])/mixcutsig[r-1],0)-
clogis((cut[int,(r)]-candcut[(r+1)])/mixcutsig[r-1],0))}
#initializes the acceptance ratio for the latent variable and cutoff parameters combined.#
ratio<-numeric()
#initializes the acceptance ratio for the beta vector#
bratio<-numeric()
#create design matrix (effects model because intercept included)#
X<-model.matrix(SEVER1~GENDER+RACE+MAR+DIS+MENT,data)
#This starts writing iterations#
#I output 19 simulations 700 times to come up with the burn of 2,000 plus the 10,000 and some
extras#
for (rep in 1:700){
#Starts calculation iterations#
for(int in 2:(length+burn)){

#Updates cutoffs and latent variables#
#Makes sure that the while loop is not bypassed#

```

```

cut[int,3]<-cut[(int-1),3]
#Simulates a candidate cutoff value from a truncated normal distribution with the upper limit being
the cutoff parameter one higher and the lower limit being the cutoff parameter one lower.#
candcut[3]<-cut[int,4]
while(round(abs(candcut[3]-(cut[(int-1),4]+cut[int,2])/2),6)>=round((cut[(int-1),4]-cut[int,2])/2,6)){
  candcut[3]<-rnorm(1,cut[(int-1),3],mixcutsig[1])
}

cut[int,4]<-cut[(int-1),4]
candcut[4]<-cut[int,5]
while(round(abs(candcut[4]-(cut[(int-1),5]+cut[int,3])/2),6)>=round((cut[(int-1),5]-cut[int,3])/2,6)){
  candcut[4]<-rnorm(1,cut[(int-1),4],mixcutsig[2])
}

cut[int,5]<-cut[(int-1),5]
candcut[5]<-cut[int,4]
while(round(candcut[5],6)<=round(cut[int,4],6)){
  candcut[5]<-rnorm(1,cut[(int-1),5],mixcutsig[3])
}

#Calculate acceptance ratio#
#Calculates the second half of the acceptance ratio found in equation 12 using the function
defined above called cpratio1#
ratio[int]<-cpratio1(2)*cpratio1(3)*cpratio1(4)
#Calculates the first half of the acceptance ratio found in equation 12#
for (i in 1:nrow(data)){
  if(data$SEVER1[i]>0){
    ratio[int]<-ratio[int]*(clogis(candcut[(data$SEVER1[i]+2)],X[i,]%*%beta[(int-1),])-
clogis(candcut[(data$SEVER1[i]+1)],X[i,]%*%beta[(int-
1),]))/(clogis(cut[int,(data$SEVER1[i]+2)],X[i,]%*%beta[(int-1),])-
clogis(cut[int,(data$SEVER1[i]+1)],X[i,]%*%beta[(int-1),]))
  }
}

#Replaces the new candidate values for the cutoff parameters with probability of the acceptance
ratio#
#Generate from a uniform [0,1] distribution to accept with the correct probability#
test<-runif(1,0,1)
if(test<ratio[int]) {
#Replaces the new candidate values for the cutoff parameters with probability of the acceptance
ratio#
cut[int,]<-candcut
#Replaces the new latent variables with the candidates simulated from the truncated logistic
distribution#
for (i in 1:nrow(data)){
  latent[1,i]<-rlogis(1,t(X[i,])%*%t(t(beta[(int-1),])),1) #replace latent[int,i] for latent[1,i] to
graph latenat variables#
  while(abs(latent[1,i]-
(cut[int,(data$SEVER1[i]+2)]+cut[int,(data$SEVER1[i]+1)]))/2)>=(cut[int,(data$SEVER1[i]+2)]-
cut[int,(data$SEVER1[i]+1)])/2){
    latent[1,i]<-rlogis(1,t(X[i,])%*%t(t(beta[(int-1),])),1)
  }
}
}

#Replaces the new betas if the candidate betas are not accepted#

```

```

beta[int,]<-beta[(int-1),]
#Generates new candidate betas from a multivariate normal distribution#
candbeta<-mvrnorm(n=1,beta[(int-1),],mixbetasig*solve(t(X)**X))
#Uses equation 14 to calculate acceptance ratio#
bratio[int]<-1
for (i in 1:nrow(data)){
bratio[int]<-
bratio[int]*logis(latent[1,i],t(X[i,])**t(t(candbeta)))/logis(latent[1,i],t(X[i,])**t(t(beta[int,]))) }
#Replaces the new betas with candidate betas with probability of the acceptance ratio#
test<-runif(1,0,1)
if(test<bratio[int]) {
beta[int,]<-candbeta
}
}
#Outputs the next 19 simulations for the parameters#
write(t(beta[2:(dim(beta)[1]),]),file="C:/Documents and
Settings/localuser/Desktop/severbeta1b.txt",ncol=28,append=TRUE)
write(t(cut[2:(dim(beta)[1]),]),file="C:/Documents and
Settings/localuser/Desktop/severcut1b.txt",ncol=6,append=TRUE)
#Stores the last simulation to generate new simulations#
beta[1,]<-beta[int,]
cut[1,]<-cut[int,]
}
#This is the R code used for my plots and credible intervals#
#Read in the data This was done for each model seperately#
data<-read.table("f:\\severbeta1.txt")
cut<-read.table("f:\\severcut1.txt")
#Take out the burn#
data<-data[2001:12000,]
cut<-cut[2001:12000,]
nrow(data)

#Plots the mixing plots (9 at a time)#
par(mfrow=c(3,3))
plot(data[,1],type="l")
title(expression(paste("Mixing Plot for Beta 0")))
plot(data[,2],type="l")
title(expression(paste("Mixing Plot for Beta 1 (Gender)")))
plot(data[,3],type="l")
title(expression(paste("Mixing Plot for Beta 2")))
plot(data[,4],type="l")
title(expression(paste("Mixing Plot for Beta 3")))
plot(data[,5],type="l")
title(expression(paste("Mixing Plot for Beta 4 (NA-White)")))
plot(data[,6],type="l")
title(expression(paste("Mixing Plot for Beta 5")))
plot(data[,7],type="l")
title(expression(paste("Mixing Plot for Beta 6")))
plot(data[,8],type="l")
title(expression(paste("Mixing Plot for Beta 7")))
plot(data[,9],type="l")
title(expression(paste("Mixing Plot for Beta 8 (Married-Never Married)")))

par(mfrow=c(3,3))
plot(data[,10],type="l")
title(expression(paste("Mixing Plot for Beta 9")))

```

```

plot(data[,11],type="l")
title(expression(paste("Mixing Plot for Beta 10")))
plot(data[,12],type="l")
title(expression(paste("Mixing Plot for Beta 11 (Austism-Retardation)")))
plot(data[,13],type="l")
title(expression(paste("Mixing Plot for Beta 12")))
plot(data[,14],type="l")
title(expression(paste("Mixing Plot for Beta 13")))
plot(data[,15],type="l")
title(expression(paste("Mixing Plot for Beta 14 (Cerebral Palsy-Retardation)")))
plot(data[,16],type="l")
title(expression(paste("Mixing Plot for Beta 15")))
plot(data[,17],type="l")
title(expression(paste("Mixing Plot for Beta 16")))
plot(data[,18],type="l")
title(expression(paste("Mixing Plot for Beta 17")))

```

```

par(mfrow=c(3,3))
plot(data[,19],type="l")
title(expression(paste("Mixing Plot for Beta 18")))
plot(data[,20],type="l")
title(expression(paste("Mixing Plot for Beta 19")))
plot(data[,21],type="l")
title(expression(paste("Mixing Plot for Beta 20")))
plot(data[,22],type="l")
title(expression(paste("Mixing Plot for Beta 21")))
plot(data[,23],type="l")
title(expression(paste("Mixing Plot for Beta 22")))
plot(data[,24],type="l")
title(expression(paste("Mixing Plot for Beta 23")))
plot(data[,25],type="l")
title(expression(paste("Mixing Plot for Beta 24")))
plot(data[,26],type="l")
title(expression(paste("Mixing Plot for Beta 25 (Profound-Mild Retardation)")))
plot(data[,27],type="l")
title(expression(paste("Mixing Plot for Beta 26")))

```

```

par(mfrow=c(2,2))
plot(data[,28],type="l")
title(expression(paste("Mixing Plot for Beta 27")))
plot(cut[,3],type="l")
title(expression(paste("Mixing Plot for Cut 1")))
plot(cut[,4],type="l")
title(expression(paste("Mixing Plot for Cut 2")))
plot(cut[,5],type="l")
title(expression(paste("Mixing Plot for Cut 3")))
#These lines were used only for frequency models#
plot(cut[,6],type="l")
title(expression(paste("Mixing Plot for Cut 4")))

```

```

#This code creates 95% credible intervals#
BUCL<-numeric()
BLCL<-numeric()
#Graph men vs. women#
var<-sort(data[,2])

```

```

BLCL[1]<-exp(var[round(.025*length(var),0)])
BUCL[1]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Women vs. Men")
abline(v=1,col="blue")
abline(v=BLCL[1],col="black")
abline(v=BUCL[1],col="black")
mean(exp(var))
#Graph black vs. white#
var<-sort(data[,3])
BLCL[2]<-exp(var[round(.025*length(var),0)])
BUCL[2]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Black vs. White")
abline(v=1,col="blue")
abline(v=BLCL[2],col="black")
abline(v=BUCL[2],col="black")

#Graph Asian vs. white#
var<-sort(data[,4])
BLCL[3]<-exp(var[round(.025*length(var),0)])
BUCL[3]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Asian vs. White")
abline(v=1,col="blue")
abline(v=BLCL[3],col="black")
abline(v=BUCL[3],col="black")

#Graph native american vs. white#
var<-sort(data[,5])
BLCL[4]<-exp(var[round(.025*length(var),0)])
BUCL[4]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Native American vs. White")
abline(v=1,col="blue")
abline(v=BLCL[4],col="black")
abline(v=BUCL[4],col="black")

#Graph other vs. white#
var<-sort(data[,6])
BLCL[5]<-exp(var[round(.025*length(var),0)])
BUCL[5]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Other vs. White")
abline(v=1,col="blue")
abline(v=BLCL[5],col="black")
abline(v=BUCL[5],col="black")

#Graph hispanic vs. white#
var<-sort(data[,7])
BLCL[6]<-exp(var[round(.025*length(var),0)])
BUCL[6]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Hispanic vs. White")
abline(v=1,col="blue")
abline(v=BLCL[6],col="black")
abline(v=BUCL[6],col="black")

#Graphs married vs. never married#
var<-sort(data[,8])
BLCL[7]<-exp(var[round(.025*length(var),0)])
BUCL[7]<-exp(var[round(.975*length(var),0)])

```



```

plot(density(exp(var)),main="Distribution of Odds Ratio for Married vs. Never Married")
abline(v=1,col="blue")
abline(v=BLCL[7],col="black")
abline(v=BUCL[7],col="black")

#Graphs separated vs. never married#
var<-sort(data[,9])
BLCL[8]<-exp(var[round(.025*length(var),0)])
BUCL[8]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Separated vs. Never Married")
abline(v=1,col="blue")
abline(v=BLCL[8],col="black")
abline(v=BUCL[8],col="black")

#Graphs divorced vs. never married#
var<-sort(data[,10])
BLCL[9]<-exp(var[round(.025*length(var),0)])
BUCL[9]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Divorced vs. Never Married")
abline(v=1,col="blue")
abline(v=BLCL[9],col="black")
abline(v=BUCL[9],col="black")

#Graphs widowed vs. never married#
var<-sort(data[,11])
BLCL[10]<-exp(var[round(.025*length(var),0)])
BUCL[10]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Widow(er) vs. Never Married")
abline(v=1,col="blue")
abline(v=BLCL[10],col="black")
abline(v=BUCL[10],col="black")

#Graphs autism vs. mental retardation#
var<-sort(data[,12])
BLCL[11]<-exp(var[round(.025*length(var),0)])
BUCL[11]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Autism vs. Mental Retardation")
abline(v=1,col="blue")
abline(v=BLCL[11],col="black")
abline(v=BUCL[11],col="black")

#Graphs blind vs. mental retardation#
var<-sort(data[,13])
BLCL[12]<-exp(var[round(.025*length(var),0)])
BUCL[12]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Blind vs. Retardation")
abline(v=1,col="blue")
abline(v=BLCL[12],col="black")
abline(v=BUCL[12],col="black")

#Graphs brain vs. retardation#
var<-sort(data[,14])
BLCL[13]<-exp(var[round(.025*length(var),0)])
BUCL[13]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Brain Damage vs. Retardation")
abline(v=1,col="blue")

```

```

abline(v=BLCL[13],col="black")
abline(v=BUCL[13],col="black")

#Graphs cerebral vs. retardation#
var<-sort(data[,15])
BLCL[14]<-exp(var[round(.025*length(var),0)])
BUCL[14]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Cerebral Palsy vs. Retardation")
abline(v=1,col="blue")
abline(v=BLCL[14],col="black")
abline(v=BUCL[14],col="black")

#Graphs chemical vs. retardation#
var<-sort(data[,16])
BLCL[15]<-exp(var[round(.025*length(var),0)])
BUCL[15]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Chemical Dependency vs. Retardation")
abline(v=1,col="blue")
abline(v=BLCL[15],col="black")
abline(v=BUCL[15],col="black")

#Graphs deafness vs. retardation#
var<-sort(data[,17])
BLCL[16]<-exp(var[round(.025*length(var),0)])
BUCL[16]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Deafness vs. Retardation")
abline(v=1,col="blue")
abline(v=BLCL[16],col="black")
abline(v=BUCL[16],col="black")

#Graphs Epilepsy vs. retardation#
var<-sort(data[,18])
BLCL[17]<-exp(var[round(.025*length(var),0)])
BUCL[17]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Epilepsy vs. Retardation")
abline(v=1,col="blue")
abline(v=BLCL[17],col="black")
abline(v=BUCL[17],col="black")

#Graphs None vs. retardation#
var<-sort(data[,19])
BLCL[18]<-exp(var[round(.025*length(var),0)])
BUCL[18]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Retardation vs. None")
abline(v=1,col="blue")
abline(v=BLCL[18],col="black")
abline(v=BUCL[18],col="black")

#Graphs physical vs. retardation#
var<-sort(data[,20])
BLCL[19]<-exp(var[round(.025*length(var),0)])
BUCL[19]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Retardation vs. Physical Health Problems")
abline(v=1,col="blue")

```

```

abline(v=BLCL[19],col="black")
abline(v=BUCL[19],col="black")

#Graphs mental vs. retardation#
var<-sort(data[,21])
BLCL[20]<-exp(var[round(.025*length(var),0)])
BUCL[20]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Mental Illness vs. Retardation")
abline(v=1,col="blue")
abline(v=BLCL[20],col="black")
abline(v=BUCL[20],col="black")

#Graphs situational vs. retardation#
var<-sort(data[,22])
BLCL[21]<-exp(var[round(.025*length(var),0)])
BUCL[21]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Situational Mental Illness vs. Retardation")
abline(v=1,col="blue")
abline(v=BLCL[21],col="black")
abline(v=BUCL[21],col="black")

#Graphs other vs. retardation#
var<-sort(data[,23])
BLCL[22]<-exp(var[round(.025*length(var),0)])
BUCL[22]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Other Disability vs. Retardation")
abline(v=1,col="blue")
abline(v=BLCL[22],col="black")
abline(v=BUCL[22],col="black")

#Graphs none vs. mild#
var<-sort(data[,24])
BLCL[23]<-exp(var[round(.025*length(var),0)])
BUCL[23]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for None vs. Mild Retardation")
abline(v=1,col="blue")
abline(v=BLCL[23],col="black")
abline(v=BUCL[23],col="black")

#Graphs moderate vs. mild#
var<-sort(data[,25])
BLCL[24]<-exp(var[round(.025*length(var),0)])
BUCL[24]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Moderate Retardation vs. Mild Retardation")
abline(v=1,col="blue")
abline(v=BLCL[24],col="black")
abline(v=BUCL[24],col="black")

#Graphs severe vs. mild#
var<-sort(data[,26])
BLCL[25]<-exp(var[round(.025*length(var),0)])
BUCL[25]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Severe Retardation vs. Mild Retardation")

```

```

abline(v=1,col="blue")
abline(v=BLCL[25],col="black")
abline(v=BUCL[25],col="black")

#Graphs profound vs. mild#
var<-sort(data[,27])
BLCL[26]<-exp(var[round(.025*length(var),0)])
BUCL[26]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Profound Retardation vs. Mild Retardation")
abline(v=1,col="blue")
abline(v=BLCL[26],col="black")
abline(v=BUCL[26],col="black")

#Graphs unknown vs. mild#
var<-sort(data[,28])
BLCL[27]<-exp(var[round(.025*length(var),0)])
BUCL[27]<-exp(var[round(.975*length(var),0)])
plot(density(exp(var)),main="Distribution of Odds Ratio for Unknown Retardation vs. Mild Retardation")
abline(v=1,col="blue")
abline(v=BLCL[27],col="black")
abline(v=BUCL[27],col="black")
#Outputs credible intervals#
out<-cbind(BLCL,BUCL)
write(t(out),file="F:/Paul's stuff/school/Thesis/CIfreq8.txt",ncol=2,append=FALSE)

```