Ordinal Response Mixed Models: A Case Study

Jade Schmidt

Department of Mathematical Sciences
Montana State University

April 24, 2012

A writing project submitted in partial fulfillment
of the requirements for the degree

Master of Science in Statistics

APPROVAL

of a writing project submitted by

Jade Schmidt

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date	Mark Greenwood Writing Project Advisor
Date	Mark C. Greenwood Writing Project Coordinator

ABSTRACT

Ordinal scale responses have always been popular in the biomedical, educational, and social science fields of study, but more recently the use of statistical methods tailored to the characteristics of ordinal responses have begun to gain in popularity. While many models have been proposed which allow the use of the ordering without treating the data as quantitative, little has been written about the assessment of the assumptions which accompany these ordinal response models. Further complicating these models and their assumption assessment are when random effects must be included to account for correlations within clusters of observations. This paper focuses on the treatment of ordinal responses, specifically focusing on ordinal response mixed models and the assumptions underlying these models. A murine model breast cancer research study was used as a case study to examine these ordinal response mixed models and methods for assessing model assumptions.

---

1 Ordinal Variables

An ordinal variable is a categorical variable whose levels have a natural ordering. In teacher evaluations, for example, students are asked to rate their instructor on a scale from poor to acceptable to excellent. A rating of poor is of course worse than a rating of acceptable, which in turn is less desirable than a rating of excellent. These Likert scales are traditionally used in social science and education research but have more recently been used in medical research as well. For example, when patients rate their symptoms or pain level, an ordinal scale is often used. In the case study discussed later, the metastis of cancer of certain organs of mice was rated on a scale of 0 to 4, with 0 meaning no metastis and a 4 meaning significant metastis in the organ.

The fact that the levels of an ordinal variable are ordered means these variables can, and for a statistical analysis must be coded as numeric. However this representation can be misleading. It should be noted that a numeric coding of an ordinal variable is simply a renaming of the group levels. In the case study example above, a 0 is simply the label associated with the group of having no metastis in the organ. It is important for two reasons to think of these codings as labels rather than values. First, even though the variable can be coded as numeric, it certainly is not continuous. There is no mouse in the case study with a metastis score of 0.5, because the only possible scores are 0, 1, 2, 3, and 4. Second, there may not be equal differences between group levels. For example, if the scale of teacher evaluations was coded numerically, poor might be a 0, acceptable a 1, and excellent coded as 2. However, the amount of effort an instructor would have to put into teaching to change a student's evaluation from a 0 to a 1 is likely to be a lot less than the amount of increased effort required to change the evaluation from a 1 to a 2, despite each change only being a one unit difference numerically. The distance between a 0 and 1 in this case is not equivalent to the distance between a 1 and 2.

2 Ordinal Response Models

There have been several methods used to analyze data in which the response variable is ordinal. If the numerically coded variable is treated as quantitative, typical least squares regression can be a simple method of analysis. However, this often used method commonly violates the assumptions of homoscedasticity as well as normality of the residuals. The predictions from such models are also difficult to interpret as the values will rarely be whole numbers and, the meaning of a predicted metastis score of 0.7 if the only values the variable can take on are 0, 1, 2, 3, or 4 is questionable. Additionally, this treatment of the ordinal response assumes the steps between levels of the variable are equal which is not always the case as demonstrated above. A second option for modeling ordinal responses would be to ignore the ordering of the variable, meaning treat the response as nominal. Multinomial logistic regression is often the choice in this instance. Again, there are problems with this analysis, most prominently the loss of information from ignoring the ordering resulting in a loss of power for the model.

Cumulative link models (CLM) are designed to handle the ordered but non-continuous nature of ordinal response data. In these models, for each level j of the ordinal response, the cumulative probability of being in level j or lower is modeled. CLM models take the following general form:

$$G^{-1}[P(Y \leq j)] = \alpha_j - X\beta.$$

In this notation, X represents the model matrix, β the vector of true coefficients for each regressor as well as the intercept, α_j the threshold for level $j, j = 1, \ldots J$ for an ordinal variable with J levels, and G^{-1} the link function. One simple way to interpret α_j and G^{-1} is by thinking of the ordinal response variable, Y, as having come from a latent, continuous variable, $Y^\star$ (Agresti, 2010, 2007, 2002). This CLM is then equivalent to an ordinary least squares regression of $Y^\star$ on the predictors, or

$$Y^\star = X\beta + \epsilon, \qquad \epsilon \sim N(0, \sigma^2).$$

In this way of thinking, α_j represent cut-off points that separate the levels of the ordinal response, or

$$Y = j \text{ if } \alpha_{j-1} < Y^\star \leq \alpha_j.$$

The link function in this situation is in fact the inverse cumulative density function of $Y^\star$. This is seen easily by applying the cumulative density function of $Y^\star$ to both sides of the CLM.

This idea is represented in Figure 1, with $Y^\star$ and Y on the y-axis and a single regressor, X on the x-axis. The straight line represents the simple linear regression of $Y^\star$ on X with the two distributions overlaid representing the distribution of the latent variable at each X value (x_1 and x_2).

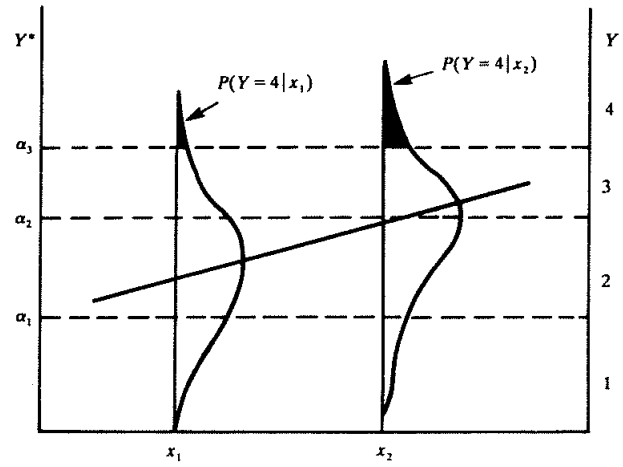

FIGURE 1: Showing the relationship between the latent, continuous variable $Y^\star$ and the ordinal variable Y. From Agresti (2010) page 54.

Figure 1 shows that shifts in X are essentially resulting in a shift in the center of the distribution of $Y^\star$. The cut-points remain the same no matter the value of X and therefore the probability of being in each category of Y changes depending on the predictor's value. Note the value of Y is solely determined by the α_j which partitions the latent continuous variable $Y^\star$.

The parameterization of subtracting $X\beta$ in the CLM above is the default parameterization in all identified programs in R (The R Core Development Team, 2011). Figure 1 above displays the logical reasoning behind this choice of parameterization. From the figure, the simple linear regression of $Y^\star$ on X has a positive slope indicating that in this case, β is greater than zero. In the CLM, this would mean that as X increases, $\alpha_j - X\beta$ will decrease, decreasing the probability of being in a lower category or, conversely, increasing the probability of being in a higher category. As can be seen in Figure 1, as X increases from x_1 to x_2, the shaded area representing the probability of Y being in category 4 also increases. The opposite is true for a negative value of β: as X increases, the probability of being in a lower category is increased.

The focus of this research was to investigate cumulative link models using R. Several packages have functions built in which model ordinal responses. For example, in the `MASS` package (Venables and Ripley, 2002), the `polr()` (proportion odds logistic regression) function takes arguments similar to a logistic model (`glm()` function), including allowing the user to set the link function. The logit, or log odds, link, which is the default, is the inverse cumulative density function of a Logistic probability distribution. When using the logit link function, CLM models are more commonly referred to as proportional odds models. Another commonly used link function is the probit link which is the inverse cumulative density function of a standard Normal distribution. The `ordinal2` package written by Christensen (2011) includes the `clm()` (cumulative link models), `clmm()` (cumulative link mixed models), and `clmm2()` functions, the first of which is similar to `polr()` while the second two allow for the addition of random effects into the CLM.

2.1 CLM Assumptions

As with all models, there are assumptions which must be satisfied in order for the results of the analysis to be valid. Independence of observations and proportional odds are the two main assumptions which pertain to CLM models. The independence assumption will be discussed further in the Section 3. The assessment of the proportional odds assumption is an important but often overlooked step in the process of model building. This assumption states that β is independent of the level j, or that the effect of X is the same for all levels j of the ordinal response. Another way of thinking about this assumption is that the difference in probit or logit of the cumulative probability for $Y \leq j$ is constant for all values of X. Notationally,

$$G^{-1}[P(Y \leq j|X)] - G^{-1}[P(Y \leq i|X)] = \alpha_j - \alpha_i.$$

Harrell (2001) and Ananth and Kleinbaum (1997) note the existence of a χ^2 score test of the proportional odds assumptions. However, information regarding the calculation of such a test statistic was not discussed and was unable to be found during further investigations. Harrell does however discuss in more depth a qualitative assessment of proportional odds. The following plot is an example of such an assessment, created using Harrell's `summary.formula()` function from his `Hmisc` package (2010) for R.

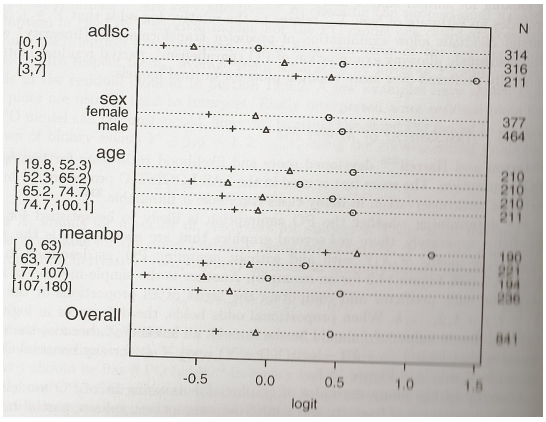

FIGURE 2: A plot assessing the proportion odds assumption for a series of both qualitative and categorical predictors. The circle, triangle, and plus sign correspond to the empirical logit of $Y \geq 1, 2, 3$, respectively from Harrell (p. 336, 2001).

In Figure 2, each symbol is the empirical logit or probit of the response variable calculated by finding the Logistic or Normal quantiles associated with the proportion of responses in the data set being less than (or greater than in the plot above) a certain category, with one symbol associated with each level of the response. The proportional odds assumption is checked by examining the vertical consistency of distances between any two of the three symbols within a variable. Note that to check this assumption for quantitative predictors, Harrell suggests binning the regressor and then plotting the probit or logit for each of the $j-1$ levels of the response for each bin. The distance between symbols represents the difference in

logit or probit values between two different values of j. If the proportional odds assumption is violated, this distance will depend on the value of X, whereas consistency across all values of the predictor indicate that the assumption is valid. However, is it difficult to tell, and Harrell does not address, how much inconsistency in distances would lead to the belief the proportional odds assumption is violated. Further research on the χ^2 score test mentioned previously would be beneficial.

It should be noted that the cumulative link model described here is in no way the only way to model ordinal responses. Agresti (2010) describes cumulative logit models which do not require the proportional odds assumption as well as adjacent categories logit models, continuation-ratio logit models, and cumulative log-log link models. These more complicated models present additional struggles in adding random effects to the model, which is already a difficult computational challenge. As mentioned previously, multinomial models could be used, but these ignore the information available in the ordinal nature of the categorical response.

3 Random Effects

The assumption of independence of observations applies to all regression models. When multiple measurements are taken on the same individual or across time, this assumption is violated. In order to account for dependent observations a random effect can be added to the previous model. Cumulative link mixed models have the following general form:

$$G^{-1}[P(Y_i \leq j)] = \alpha_j - (Z_{t[i]}u_t + X_i\beta)$$

$$\text{where } u_t \sim N(0, \sigma_u^2).$$

In this notation, u_t represents the vector of coefficients corresponding to the group-level predictors $Z_t[i]$ for observation i in cluster t. This model has the added assumption that the random effects are Normally distributed and centered at zero. The random effect induces the correlation expected between observations in the same cluster and allows inferences to be made to the population from which the groups were sampled. It should be noted that model estimates can be unstable if there are a small number of observations within clusters or if there are few clusters from which to estimate within group correlation.

Adding random effects to an ordinal response model will further complicate an already complex likelihood for the observations. As coefficient estimates and standard errors are calculated using maximum likelihood methods, the result is often difficulty in model convergence. Assuming an underlying Logistic distribution can make convergence issues even more likely. A logit-link model with random effects will create a mixed likelihood that combines the Normal distribution of the random effect with the Logistic distribution assumed for the latent responses. Using a probit-link, which assumes a Normal latent distribution for the data, can help with model convergence. Currently, the `clmm()` function in the `ordinal2` package (Christensen, 2011) uses Laplace approximations to fit the model. However, it is soon expected to be able to estimate model parameters using either standard or adaptive

Gauss-Hermite quadrature approximation, an option currently available in the `clmm2()` function. However, the `clmm2()` function is limited to only accepting a single random effect. See Agresti (2010) or Hedeker and Gibbons (1994) for more information on estimation of ordinal response mixed models.

When a random effect is included in a model, it is important to look at the intra-class correlation (ICC). ICC is defined as the correlation of observations within a group and is a way to look at how similar these within-cluster observations are to one another. The following formula is used to calculate ICC:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}.$$

Here, σ^2 represents the residual variance and in the case of CLM models, is assumed to be one while σ_u^2 represents the variance of the random effect. Values of ICC near one indicate that observations within a cluster are very similar to one another, while values close to zero indicate that the random effect may not be necessary as observations within a group are nearly independent. For the probit cumulative link mixed model, the residual variance is the variance of the latent response and therefore is one by definition of the standard Normal distribution.

4 Case Study

A dataset investigating the effect of bio-energy treatments on breast cancer in a murine model is used to illustrate these methods. In this experiment, male mice were injected with breast cancer and treated for 15 days. Five different treatments were investigated: healing touch administered three days per week (IIH), healing touch administered daily (IH), reiki administered three times per week (IIR), reiki administered daily (IR) and a control group. Reiki and healing touch are both bio-energy treatments in which the healer uses hand placements and thoughts to aid the flow of energy throughout the patients body. The hope is that this flow of energy will help the body heal itself. Reiki is an ancient Japanese treatment with the theory passed down in a master-apprentice relationship. The goal of this treatment is to help the natural flow of energy through the body. Healing touch on the other hand is more modern with techniques being taught in schools around the globe. This treatment focuses on directing the movement of energy through the patient's body (Potter, 2003). Both of these treatments have vast amounts of anecdotal evidence with human patients. However, the physical improvements of patients is often attributed to the placebo effect by critics of bio-energy treatments. The application of these treatments to mice removes the possibility of the placebo effect as an explanation. Seven research mice were placed in a cage, with at least two cages per treatment. During an application of the bio-energy treatments, all cages getting the same treatment were placed on a table and treated simultaneously. After the 15 day treatment period, the mice were euthanized and dissected. Samples which included tumor and various organs were sent to a pathologist to grade the rate of metastis. Metastis was scored on a ordinal scale from 0 to 4, with a 0 indicating no metastis, or healthy organ

cells, and a 4 indicating significant metastis, or extremely cancer filled organ cells. Note that this was not a balanced design with the control group occupying six cages, reiki treatments being applied to four cages each and healing touch treatments only getting two cages each. Due to the way the cancer spreads, the lungs, liver, and spleen were the organs of most interest in this study.

4.1 Data and Model

Occasionally when organ or tumor samples were taken from the mice, multiple organs or those not targeted were sampled as well. Therefore, there were some mice with multiple metastis scores on a single organ. In this case, the maximum metastis score was taken as this was felt to be a better indication of whether the cancer had spread than the minimum metastis score. Now with one measurement per mouse per organ, a separate analysis was conducted for each organ studied: the liver, lungs, and spleen. Two different analyses were conducted on each organ. Although the mice are genetically identical and cages treated as similarly as possible, the social nature of mice leads to the belief that mice within a cage will be more similar to one another than to mice in another cage receiving the same treatment. This correlation amongst mice within a cage is accounted for through the use of a random intercept for each cage in the first analysis which has the following form:

$$\Phi^{-1}[P(Y_i \leq j)] = \alpha_j - (\mathbf{1}u_{c[i]} + X_i\beta) \text{ where } u_c \sim N(0, \sigma_c^2).$$

Here, Φ^{-1} represents the inverse cumulative density function of a standard Normal distribution, or more simply the probit link, $u_{c[i]}$ is a vector of random intercept coefficients for the cage c where mouse i was housed, X_i is a the model matrix which includes an intercept which represents the baseline treatment (control) and indicator variables which represent deviations from the baseline for each treatment other than the control, and β represents the coefficient vector for the control group and deviations from the control group for each other treatment. Since, at most, only 7 measurements were available for each cage, there was concern the model estimates from this analysis would be unstable. Of further concern was the fact that the effect of cage may be hiding a treatment effect as treatments were applied to cages as a whole and so these two variables are somewhat confounded. The second analysis assumed there was no effect of cage, or that each mice was independent. This model had the following form:

$$\Phi^{-1}[P(Y_i \leq j)] = \alpha_j - X_i\beta$$

with Φ, X, and β having the same meaning as the previous model. It should be noted that because treatments were applied to all cages receiving the same treatment at the same time, the experimental units are actually at the treatment level rather than the cage level.

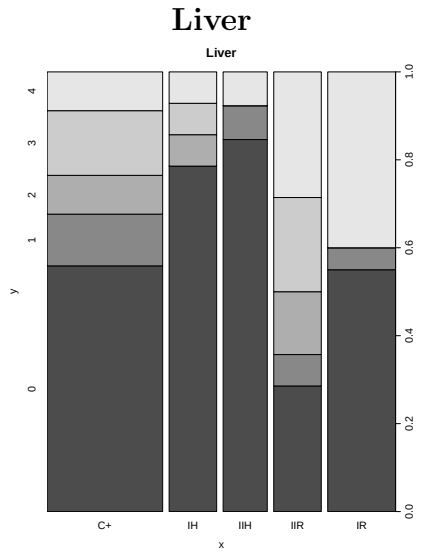

FIGURE 3: A mosaic plot showing the observed proportions of mice with liver measurements in each metastis score level for each treatment. The treatments along the x-axis are, in order, the control group (C^+), daily healing touch (IH), three days per week healing touch (IIH), three days per week reiki (IIR) and daily reiki (IR).

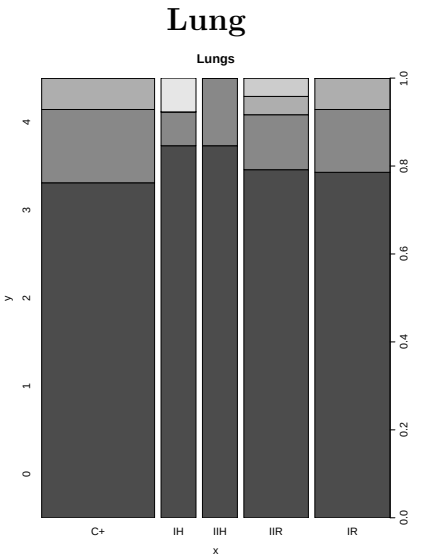

FIGURE 4: A mosaic plot showing the observed proportions of mice with lung measurements metastis score level for each treatment.

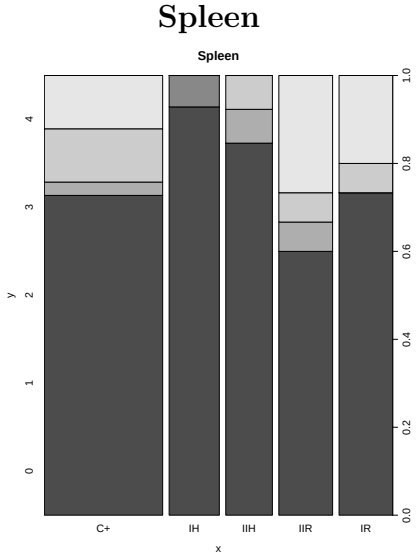

FIGURE 5: A mosaic plot showing the observed proportions of mice with spleen measurements in each metastis score level for each treatment.

Figures 3, 4, and 5 display the responses for each treatment within each organ: the liver, lung, and spleen, respectively. From these plots, across all organs the healing touch groups showed more mice with lower metastis scores that the control, indicating a healthier mouse, although the difference seems most apparent in the spleen and least apparent in the lungs.

The reiki treatment on the other hand shows approximately equivalent proportions of mice in each metastis level in the lungs but more mice with higher liver and spleen metastis scores when compared to the control group. This indicates that reiki may in fact be helping the cancer spread to the latter two organs.

4.2 Results and Diagnostics

4.2.1 Liver Organ

In the liver data, the two analyses have similar results with treatment coefficients having the same sign and being within one standard error of the each other. However the p-values for the coefficients are significantly lower in the non-mixed model. The coefficient estimates for the healing touch group are both negative, indicating that in comparison to the control group, both healing touch groups appeared to have higher probabilities of remaining in lower metastis scores, indicating a healthier mouse. The opposite was true for the reiki groups, with positive coefficients indicating higher probabilities of being in higher metastis scores, or sicker mice. Tables 1 and 2 give the model coefficients, standard errors, Z-test statistics for whether the coefficients are zero or not and p-values for this test for the mixed model and the model with no random effects, respectively.

	Estimate	Std. Error	z value	Pr(>|z|)
0|1	0.12	0.37	0.32	0.75
1|2	0.37	0.37	1.00	0.32
2|3	0.62	0.37	1.64	0.10
3|4	1.06	0.39	2.74	0.01
GroupIH	-0.64	0.74	-0.87	0.38
GroupIIH	-0.88	0.76	-1.16	0.25
GroupIIR	0.71	0.66	1.08	0.28
GroupIR	0.52	0.61	0.86	0.39

TABLE 1: Model estimates for the liver data for the model which includes cage as a random effect.

	Estimate	Std. Error	z value	Pr(>|z|)
0|1	0.37	0.33	1.14	0.25
1|2	0.73	0.33	2.19	0.03
2|3	1.06	0.34	3.08	0.00
3|4	1.65	0.38	4.38	0.00
GroupIH	-0.88	0.72	-1.22	0.22
GroupIIH	-1.34	0.83	-1.60	0.11
GroupIIR	1.05	0.57	1.85	0.06
GroupIR	0.56	0.56	0.99	0.32

TABLE 2: Model estimates for the liver data for the model which includes no random effects.

In the cumulative link mixed model, a likelihood ratio test comparing a model with no treatment effect to the model described above gave a χ^2 test statistic of 4.8493 on 4 degrees of freedom resulting in a p-value of 0.3031. This relative high value indicates there is not

a significant difference between the two models, or that there is no evidence of a treatment effect. For the liver, it indicates that treatment was not a significant predictor of metastis score. The variance of the cage random effect was estimated to be 0.538, giving an intraclass correlation between metastis scores for mice in the same cage of 0.35. However, in the second method of analysis which does not take into account the possible cage effect, the χ^2 test of a treatment effect has a test statistic of 12.300 on 4 degrees of freedom for a p-value of 0.0153. This indicates that treatment is a significant predictor of metastis score in the liver, directly contradicting the first method of analysis. A pairwise comparison of all treatments was conducted after this finding of significant differences between treatments. Bonferroni-corrected p-values for all pairwise treatment comparisons were found. With five treatments, this is 10 comparisons, so the p-values for each comparison were multiplied by 10 to get the Bonferroni-corrected p-value, a very conservative way to test equivalence of treatments. The following pairwise comparisons had the lowest Bonferroni-corrected p-values: daily healing touch and three-times-per-week reiki (0.16) as well as three-times-per-week healing touch and three-times-per-week reiki (0.08).

Figures 5 and 6 show the fitted probabilities of being in each metastis level for each treatment for the mixed model holding the random effect at zero (left) and the non-mixed model (right).

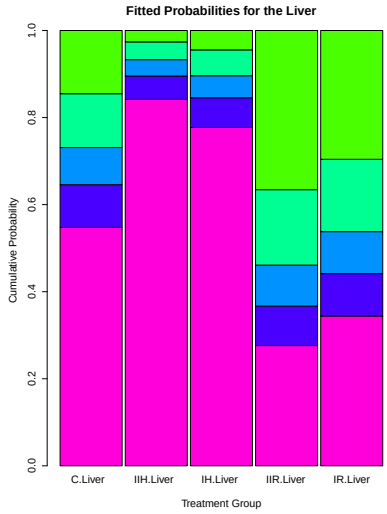

FIGURE 6: Fitted probabilities for each treatment in the liver from the cumulative link mixed model. Note in this plot the random effects for cage are set to be zero for every cage.

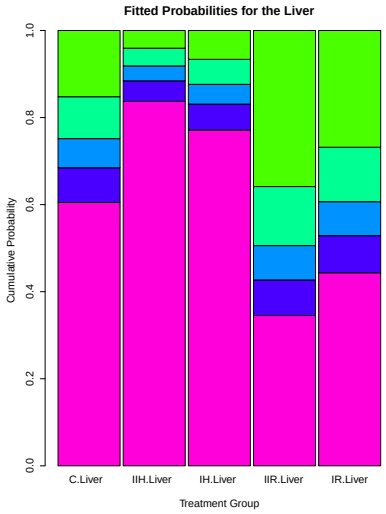

FIGURE 7: Fitted probabilities for each treatment in the liver from the second analysis.

Figure 6, which sets all random effects to be zero, and Figure 7 are similar to each other and to Figure 3 indicating the model appears to be generating fitted probabilities which are similar to those observed in the data. It seems that the model with no random effects predicts more low scores in the healing touch groups and more high scores in the control and both reiki groups than the mixed model does, which accounts for the significant differences

between treatments found on the second analysis. The plots below are assessing model assumptions using Harrell's plot to assess proportional odds (Fig. 8) and a Normal QQ plot to assess normality and magnitude of the random cage effects (Fig. 9). The distance between any two sets of points does not appear to be consistent in Figure 8, indicating there may be a violation of the proportional odds assumption. It is important to note that Harrell's (2001) plot does not allow for the inclusion of random effects which could make the difference in making probits appear more consistent across the levels of treatment. Figure 9 does indicate the assumption of normality of the random effect is valid. The variability of the random effect between -1 and 1.2 on the latent standard normal scale demonstrates the importance of the random effect in this model.

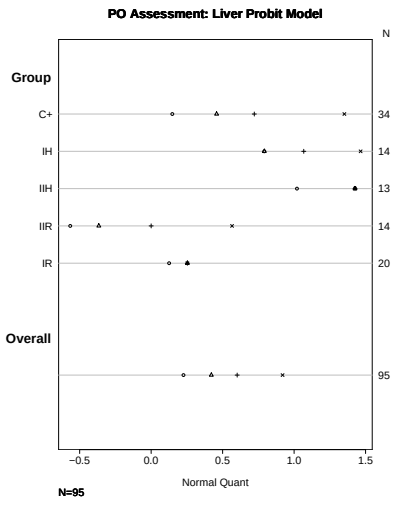

FIGURE 8: Plot of assess the proportion odds assumption for the liver. Note this plot does not take into account the random effect for cage which should be included in the model.

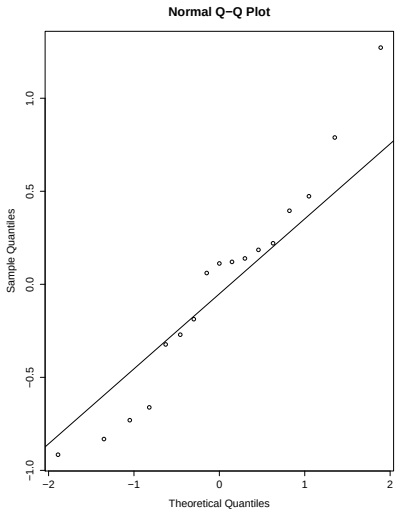

FIGURE 9: Normal QQ Plot of the random effects for cage in the model which includes treatment. The ordered random effects appear to coincide with the Normal quantiles giving no reason to feel the Normality assumption of the random effect is violated.

4.2.2 Lung Organ

In the lung data, the two analyses again give similar estimated coefficients with all being within one standard error in the two analyses. Like the liver data, the coefficient estimates for the healing touch group are both negative, indicating that in comparison to the control group, both healing touch groups appeared to have higher probabilities of remaining in lower metastis scores, or a healthier mouse. The same was true for both reiki groups, with negative coefficients indicating higher probabilities of being in lower categories. Tables 3 and 4 on the following page give the model coefficients, standard errors, Z-test statistics for whether the coefficients are zero or not and p-values for this test for the mixed model and the model with no random effects, respectively. Note that in the mixed model, standard errors for the first three thresholds were infinite. This may be the result of a two factors. First, the

estimated variance of the random effect is very close to zero with may have caused a problem with the estimability of the variance-covariance matrix. Second, looking at Figure 4, very few observations in any of the treatment groups had high level metastis scores, with the control, daily healing touch, three days per week healing touch and daily reiki devoid of any observations for at least one level of the response. This too may have caused an problem with estimability of the thresholds.

	Estimate	Std. Error	z value	Pr(>|z|)
0|1	0.75			
1|2	1.44			
2|3	2.07			
3|4	2.34	0.26	9.03	0.00
GroupIH	-0.08	0.39	-0.21	0.83
GroupIIH	-0.35	0.42	-0.83	0.40
GroupIIR	-0.03	0.28	-0.10	0.92
GroupIR	-0.06	0.26	-0.24	0.81

TABLE 3: Model estimates for the lung data for the model which includes cage as a random effect.

	Estimate	Std. Error	z value	Pr(>|z|)
0|1	1.18	0.36	3.28	0.00
1|2	2.49	0.46	5.44	0.00
2|3	3.93	0.76	5.14	0.00
3|4	4.63	1.04	4.45	0.00
GroupIH	-0.46	0.85	-0.55	0.58
GroupIIH	-0.58	0.84	-0.69	0.49
GroupIIR	-0.13	0.62	-0.21	0.83
GroupIR	-0.12	0.58	-0.21	0.83

TABLE 4: Model estimates for the lung data for the model which includes no random effects.

A likelihood ratio test comparing mixed model with no treatment effect to the mixed model which includes treatment as a predictor gave a χ^2 test statistic of 0.5941 on 4 degrees of freedom resulting in a p-value of 0.9637. This indicates there is no significant difference between the two models, or treatment was not a significant predictor of metastis score in the lung. For the lungs, the variance of the random effect was found to be 7.4×10^{-8} which made the ICC between metastis scores for mice in the same cage nearly zero. This indicates there may be no need for a random effect for cage when analyzing the lung data. In the second analysis which does not include a cage random effect, the comparison of the no treatment effect model to the model with treatment calculated the χ^2 test statistic to be 0.6042 on 4 degrees of freedom for a p-value of 0.9626, a very similar result to that found when the random effect for cage was included in the model. Figures 10, 11, and 12 on the following page show the fitted probabilities of being in each metastis level for each treatment for the model which includes the random effect for cage with the random effect set to zero (Fig. 10) and the model with no random effects (Fig. 11) as well as the proportional odds assessment using Harrell's figure (Fig. 12). Figure 10, which does set all random effects to be zero, and Figure 11 are nearly identical and both are similar to Figure 4 indicating the model

appears to be generating fitted probabilities which are similar to those observed in the data. However, the distance between any two sets of points does not appear to be consistent in Figure 12, indicating there may be a violation of the proportional odds assumption. Since the random effect does not appear necessary in the lung organ, it is likely a proportion odds model is not valid for these data. Plotting the random effects for cage, all appear to be zero indicating again that there may be no need for a random effect for the responses. This uninformative plot was not included.

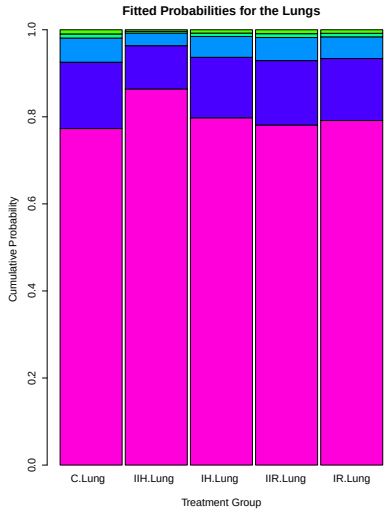

FIGURE 10: Fitted probabilities for each treatment in the lungs for the model which includes cage as a random effect. Note in this plot the random effects for cage are set to be zero for every cage.

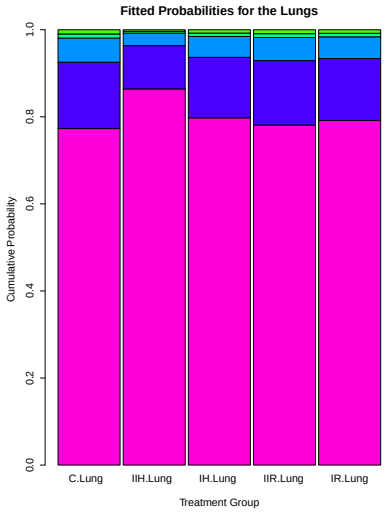

FIGURE 11: Fitted probabilities for each treatment in the lungs for the model with no random effects.

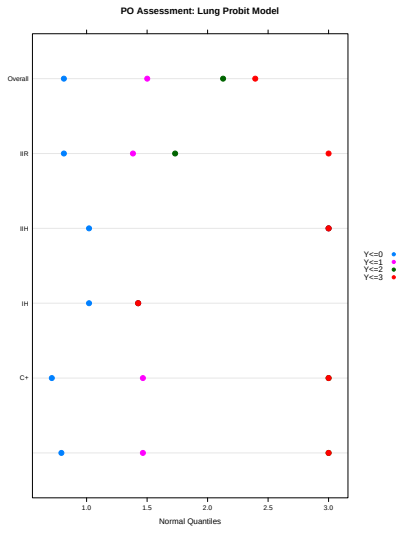

FIGURE 12: Plot of assess the proportion odds assumption for the lung. Note this plot does not take into account the random effect for cage which should be included in the model.

4.2.3 Spleen Organ

In the spleen data, the two analyses again give similar estimated coefficients with all being within one standard error in the two analyses. Like the liver and lung data, the coefficient estimates for the healing touch group are both negative, indicating that in comparison to the control group, both healing touch groups appeared to have higher probabilities of remaining in lower metastis scores, or a healthier mouse. Like the liver data, the opposite was true for both reiki groups, with positive coefficients indicating higher probabilities of being in higher categories. Tables 5 and 6 on the following page give the model coefficients, standard errors, Z-test statistics for whether the coefficients are zero or not and p-values for this test for the mixed model and the model with no random effects, respectively.

	Estimate	Std. Error	z value	Pr(>|z|)
0|1	0.55	0.59	0.93	0.35
1|2	0.61	0.59	1.02	0.31
2|3	0.80	0.60	1.34	0.18
3|4	1.36	0.62	2.20	0.03
GroupIH	-1.55	1.31	-1.18	0.24
GroupIIH	-1.03	1.25	-0.82	0.41
GroupIIR	1.07	1.01	1.06	0.29
GroupIR	0.38	0.97	0.40	0.69

TABLE 5: Model estimates for the spleen data for the model which includes cage as a random effect.

	Estimate	Std. Error	z value	Pr(>|z|)
0|1	0.98	0.39	2.52	0.01
1|2	1.04	0.39	2.67	0.01
2|3	1.25	0.40	3.14	0.00
3|4	1.87	0.44	4.25	0.00
GroupIH	-1.66	1.11	-1.50	0.13
GroupIIH	-0.81	0.85	-0.95	0.34
GroupIIR	0.64	0.64	1.00	0.32
GroupIR	0.08	0.70	0.12	0.91

TABLE 6: Model estimates for the spleen data for the model which includes no random effects.

For the model which included cage as a random effect, likelihood ratio test comparing the no treatment effect model with the model described above gave a χ^2 test statistic of 4.9568 on 4 degrees of freedom resulting in a p-value of 0.2918. This relative high value indicates that treatment was not a significant predictor of metastis score in the spleen. The variance of the random effect for cage in the mixed model was estimated to be 1.50 giving an intraclass correlation between metastis scores for mice in the same cage of 0.68. This is the highest of the three organ analyses. For the model with no random effects,a χ^2 test for the treatment

effect gave a test statistic of 7.5739 on 4 degrees of freedom for a p-value of 0.1085. Treatment may be a significant predictor of metastis score in the spleen in the non-mixed model. However, only daily healing touch and three-times-per-week reiki treatment comparison gave a Bonferroni-corrected p-value that was close to significant (0.160).

Figures 13 and 14 below show the fitted probabilities of being in each metastis level for each treatment for the mixed model (Fig. 13) and the non-mixed model (Fig 14). Figure 13, which does set all random effects to be zero, and Figure 14 are similar and both resemble Figure 5 indicating the model appears to be generating fitted probabilities which are similar to those observed in the data. Like in the analysis of the liver data, the model which does not include random effects appears to increase the estimated probabilities of being in lower categories in the healing touch groups while increasing the expected probability of being in higher categories in the control and both reiki groups. These differences account for the different ANOVA results described above.

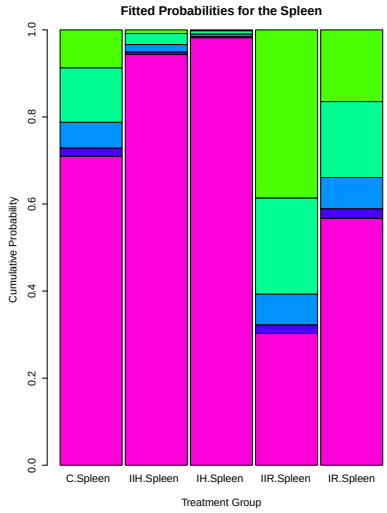

FIGURE 13: Fitted probabilities for each treatment in the spleen for the mixed model. Note in this plot the random effects for cage are set to be zero for every cage.

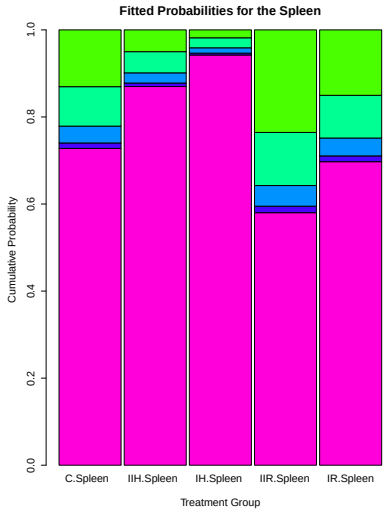

FIGURE 14: Fitted probabilities for each treatment in the spleen for the model which does not include cage as a random effect.

Figures 15 and 16 on the following page assess model assumptions with Harrell's proportional odds plot (Fig. 15) and a Normal QQ plot to assess the normality of the random cage effect for the mixed model (Fig. 16). Again we see the distance between any two sets of points does not appear to be consistent in Figure 15, indicating there may be a violation of the proportional odds assumption. It is important to note that Harrell's (2001) plot does not allow for the inclusion of random effects which could make the difference in probits appear more consistent across the levels of treatment, especially considering the high intraclass correlation seen in the spleen data. Figure 16 shows some deviation from Normality in the upper tail for the random effect of cage.

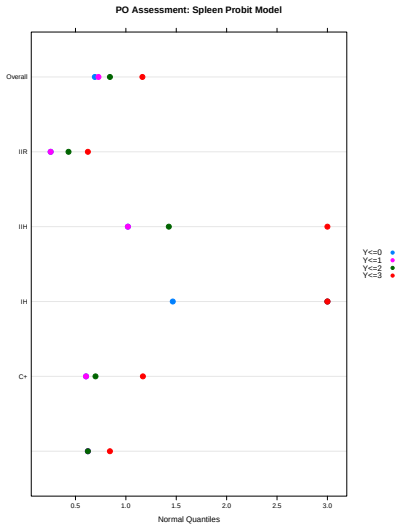

FIGURE 15: Plot of assess the proportion odds assumption for the spleen. Note this plot does not take into account the random effect for cage which should be included in the model.

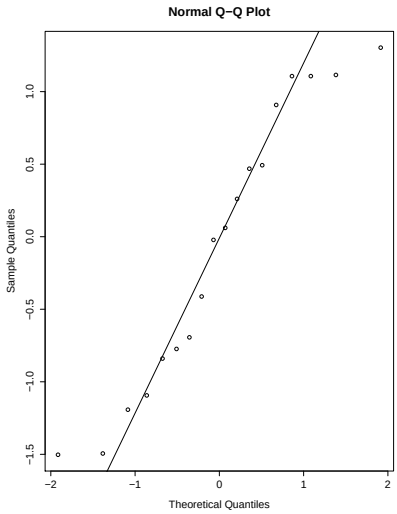

FIGURE 16: Normal QQ Plot of the random effects for cage in the model which includes treatment. The ordered random effects appear to coincide with the Normal quantiles giving no reason to feel the Normality assumption of the random effect is violated.

The fact that for two of the three organs, the results of the two analyses are not similar is concerning. Further investigation of the dataset will be necessary to determine which analysis is more appropriate.

4.3 Alternative Analyses

A logit-link could be substituted for the probit-link used in the ordinal mixed model analysis described above. The logit analysis of the lung and spleen data is nearly identical to the results of the probit model. A likelihood ratio test comparing an intercept only model to a model which includes treatment as a predictor gave a χ^2 value for the lungs of 0.6675 (compared to 0.5941 using the probit link) and for the spleen of 4.8782 (compared to 4.9568 using the probit link). The results of the likelihood ratio test for the lungs is slightly different between the two link functions (χ^2 test statistic of 4.8493 for the probit-link, p-value of 0.3031, χ^2 test statistic of 5.5253 for the logit-link, p-value of 0.2375), but both models indicated that treatment is not a significant predictor of metastis score. Additionally, although the coefficient estimates and standard errors differ between the logit and probit models, the p-values of the coefficients are very similar for all three organs.

Another mixed modeling approach that could be used would be to fit nested random effects, one for cage and one for mouse within cage. This approach would require the individual datasets from each organ to be combined into a single dataset. The cumulative link mixed model fit would then include treatment, organ, and the interaction between these two variables as well as two nested random effects, for cage as well as for mouse within cage.

5 Conclusions

Many possible methods of analysis exist for ordinal response data. If the levels of the ordinal response possibly may not be equally spaced, then ordinary least squares regression is likely not an appropriate analysis. Figures 14 and 15 are diagnostic plots of the case study data for the spleen from a linear mixed model. This plot demonstrates the common problem when ordinal responses are treated as quantitative. The Normal QQ plot (figure 15) displays extreme violations of the assumption of Normal residuals and the homogeneity of variance assumption may also be violated based on the football-shaped pattern in the Residuals vs. Fitted plot (figure 14).

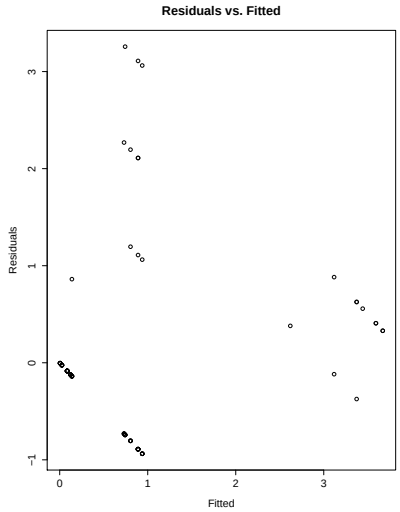

FIGURE 14: Residuals vs Fitted plot for a linear mixed model including cage as a random effect for the spleen data.

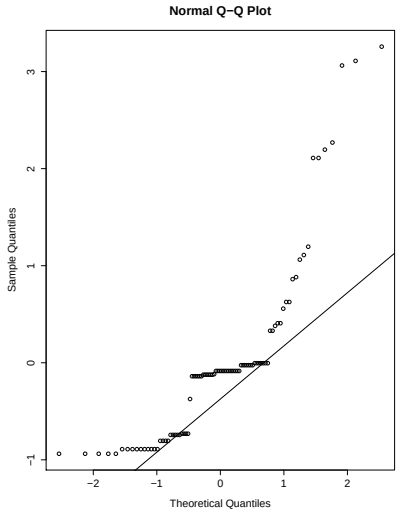

FIGURE 15: Normal QQ plot for a linear mixes model including cage as a random effect for the spleen data..

There is also the option to ignore the natural ordering of the levels and treat an ordinal response as nominal. However, not only would the nominal response analysis have less power, ordinal response models are simpler to interpret with a greater variety of modeling options including easy access to ordinal mixed models. The CLM described here is the most common ordinal response model and at a minimum improves upon nominal and quantitative analyses because the response is treated correctly.

The proportional odds assumption of cumulative link models states that the effect of the predictors should be independent of the level j of the response. The ability to assess this assumption was initially a focus of this paper but a lack of assessment tools stymied the investigation. Harrell (2001) and Ananth and Kleinbaum (1997) mention the existence of a χ^2 Score test of the proportional odds assumption but do not give any details on how to conduct the test or an interpretation of the results. Harrell (2001) does provide a qualitative assessment of this assumption via the plots seen throughout this paper. These plots however do not give a definitive answer as to whether the proportional odds assumption is violated or

not. Additionally, these plots are not able to take into account random effects and therefore may not even be a valid assessment for mixed models. In the future, it would be helpful to research this Score test more and to try to adjust Harrell's `Hmisc` package so that the `summary.formula()` function can account for a mixed model.

Finally, with regard to this case study, the sample sizes were relatively small for some treatments. For example, only 14 mice were subjected to the daily healing touch treatment. In the liver and spleen, the plots of the data and fitted probabilities appear to show differences in treatment which indicate reiki may in fact increase the spread of cancer while healing touch may have helped the mice hold the cancer at bay. The fact that no significant treatment effect was found for any of the three organs may be the result of such small sample sizes. This result could also be the product of an unstable model due to few observations within a cage. Continued research in this area could repeat this experiment with a larger sample size. The amount of training and practice of the people administering the treatments could also be of interest in further bio-energy research.

6 Acknowledgments

I would like to thank Dr. Mark Greenwood of the Mathematical Sciences Department at Montana State University for giving significant input on modeling in R as well as comments on improving this paper. I would also like to thank Dr. Megan Higgs and Dr. Jim Robison-Cox of the same department for additional help creating plots in R. And finally thank you to Dr. Alice Running and her research team for allowing me to use her study in this exploration of ordinal mixed models.

References

Agresti, Alan (2010). *Analysis of Ordinal Categorical Data (2nd Ed).* John Wiley & Sons, Inc. Hoboken, NJ.

Agresti, Alan (2007). *An Introduction to Categorical Data Analysis (2nd Ed).* John Wiley & Sons, Inc. Hoboken, NJ.

Agresti, Alan (2002). *Categorical Data Analysis (2nd Ed).* John Wiley & Sons, Inc. Hoboken, NJ.

Ananth, Cande and David Kleinbaum (1997). Regression Models for Ordinal Responses: A Review of Methods and Applications. *International Journal of Epidemiology* 26, 1323-1333.

Christensen, R. H. B. (2011). ordinal2—Regression Models for Ordinal Data R package version 2011.05-10 `http://www.cran.r-project.org/package=ordinal/`

Fox, John (2008). *Applied Regression Analysis and Generalized Linear Models (2nd Ed).* SAGE Publications. Thousand Oaks, CA

Harrell, Frank E. Jr <f.harrell@vanderbilt.edu> and with contributions from many other users. (2010). Hmisc: Harrell Miscellaneous. R package version 3.8-3. http://CRAN.R-project.org/package=Hmisc

Harrell, Frank E. Jr. (2001). *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis.* Springer Science+Business Media, Inc. New York, NY.

Hedeker, Donald and Robert D. Gibbons (1994). A Random-Effects Ordinal Regression Model for Multilevel Analysis. *Biometrics* vol. 50, no. 4, 933-944.

Pinheiro, Jose; Bates, Douglas; DebRoy, Saikat; Sarkar, Deepayan and the R Development Core Team (2011). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-101.

Potter, Pamela. What are the Distinctions Between Reiki and Therapeutic Touch?. *Clinical Journal of Oncology Nursing* vol. 7, no. 1, 89-91.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Simonoff, Jeffrey S. (2003). *Analyzing Categorical Data.* Springer-Verlag New York, Inc. New York, NY.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S (4th Ed).* Springer, New York. ISBN 0-387-95457-0