Approaches for Missing Data in Ordinal Multinomial Models

Niloofar Ramezani[1]
[1]Department of Applied Statistics and Research Methods, University of Northern Colorado, Greeley, CO, 80639

Abstract
Most of the existing research about the choice of missing data method for non-normal data has been carried out using binary data. This study however uses ordinal data to compare the different approaches of listwise deletion, mean imputation, and multiple imputation to determine how informative each method will be within an ordinal multinomial logistic model. Imputing categorical variables which are non-normal is challenging and it still is unclear which approach should be preferred (Lee et al., 2012). Considering the type of missing data (MCAR, MAR, or MNAR) is also important in determining how to handle missing values. In this study, after learning about the type of missingness by applying a logistic regression, an ordinal multinomial logistic regression is fitted to the ordinal data and within that model, different approaches of missing data are performed to evaluate the appropriateness of missing data handling procedures. This comparison is done by applying these methods to a dataset on the length of stay for people with severe mental illness at a live-in healing community in North Carolina, which includes longitudinal ordinal and multinomial data containing missing values.

Key Words: Ordinal data, Ordinal multinomial logistic regression, Imputation, Missing data, Longitudinal data

1. Introduction

Modeling length of stay (LOS) using ordered groups requires applying ordinal models due to ranking of data. When modeling LOS for individuals with persistent psychological health conditions at a live-in healing community in North Carolina which included missing values, there is a need for appropriate models of ordinal data with missing observations. It is important to predict LOS to allocate appropriate funding to this community.

Ordinal logistic regression models have been applied in recent years in analyzing data with ranked multiple response outcomes. Ordered information has been increasingly used in health indicators but their use in the public health is still rare (Abreu et al., 2009). This may be attributed to these models' complexity, assumptions validation, and limitations of modeling options offered by statistical packages (Lall, 2002). Missing values that are present when dealing with real data most of the time add more complexity to ordinal models, but not much research exists about their handling techniques especially within ordinal models. Regardless of their complexity, ordinal hypothesis tests provide increased power and ordinal logistic models allow for interpretations based on inherent rankings; therefore, increased accessibility of these models is important, particularly choosing among link functions such as cumulative logits, adjacent-category logits, and continuation-ratio logits, and choosing between missing data approaches like listwise deletion and imputation methods.

Using the mental health data, this study compares different link functions within ordinal multinomial logistic models and evaluates the appropriateness of missing data methods using SAS (The SAS Institute, Cary, NC) and R (R Development Core Team, 2015). The variables affecting the length of stay in the healing community, such as race, gender, and health conditions, are also presented and their significance and effect on the LOS response are evaluated and used as a way of comparing different models and procedures within the two different statistical software used in this study.

2. Ordinal Multinomial Logistic Regression

The multinomial logistic regression model is an extension of the binomial logistic regression model. This type of model is used when the dependent variable has more than two nominal (unordered) categories. When the response categories are ordered, a multinomial regression model still can be used. According to Agresti (2007), the disadvantage is that some information about the ordering is thrown away. An ordinal logistic regression model preserves that information, but it is slightly more involved which is the model that is used in this study.

There are different logit functions such as Cumulative Logit, Adjacent–Categories Logit, and Continuation Ratio Logit which are used within regression models to provide useful extensions of the multinomial logistic model to ordinal response data. The author proposes fitting these models in the presence of missing data in this paper. Each of these models are briefly explained below according to the notations used in Agresti (2013):

2.1 Cumulative Logit Models

The cumulative logit function used in ordinal multinomial logistic models is as below modeling categories $\leq \boldsymbol{j}$ versus categories $> \boldsymbol{j}$

$$logit\ (P(Y \leq j)) = log\left(\frac{P(Y \leq j)}{P(Y > j)}\right) = log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right)$$
$$= log\left(\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right),\ for\ j = 1, \cdots, J-1$$

Using this logit function, the cumulative logit model is as below

$$logit\ (P(Y \leq j)) = \alpha_j + \sum_{k=1}^{K} \beta_k X_k.$$

2.2 Adjacent–Categories Logit Models

The adjacent-categories logit function used in ordinal multinomial logistic models is as below modeling two adjacent categories

$$log\left(\frac{P(Y = j)}{P(Y = j+1)}\right) = log\left(\frac{\pi_j}{\pi_{j+1}}\right).$$

Using this logit function, the adjacent-categories logit model is as below

$$log\left(\frac{\pi_j}{\pi_{j+1}}\right) = \alpha_j + \sum_{k=1}^{K} \beta_k X_k.$$

Within this model, only adjacent categories will be used in odds resulting in using local odds ratios for interpretations, whereas within the cumulative logit models, the entire response scale is used for the model and cumulative odds ratio is used for their interpretation.

2.3 Continuation–ratio Logit

The continuation-ratio logit function used in ordinal multinomial logistic models is as below

$$logit\ \left(\omega_j(X)\right) = log\left(\frac{P(Y=j)}{P(Y \geq j+1)}\right)$$
$$= log\left(\frac{\pi_j}{\pi_{j+1} + \cdots + \pi_J}\right),\ for\ j = 1, \cdots, J-1$$

where $\omega_j(X) = \frac{\pi_j(X)}{\pi_j(X) + \cdots + \pi_J(X)}$.

Using this logit function, the continuation-ratio logit model is as below

$$logit\ (\omega_j(X)) = \alpha_j + \sum_{k=1}^{K} \beta_k X_k.$$

As described in Agresti (2013), this model is useful when a sequential mechanism determines the response outcome. Mechanisms like survival through various age periods would be suitable for such models.

3. Missing Data Handling Techniques

Missing data presents a challenge in any type of research including this study within ordinal models. Missing data is associated with numerous statistical concerns (Cheema, 2014), and the severity of the problem depends largely on the type (Rubin, 1976) and quantity (Gibson & Olejnik, 2003) of missing data. Various missing data handling procedures are available to researchers, but the procedures vary in regards to overall effectiveness and technical skill required for implementation (Gibson & Olejnik, 2003).

There are different missing data handling techniques such as listwise deletion, mean imputation, and multiple imputation that are being discussed in this study. There are other methods of handling missing data such as maximum likelihood using the EM algorithm that are not addressed in this paper. Below are the brief descriptions of each of these three missing data handling methods:

3.1 Listwise Deletion

Using this strategy, any individual in a data set is deleted from an analysis if there are missing data on any variable in the analysis. In a review, Cheema (2014) found listwise deletion to be the most common handling procedure used in educational research and also in other fields. Listwise deletion is easy to use and is often the default in statistical packages, but it can be leaded to a dramatic loss in power, especially if missing values are distributed across several variables (Schafer & Olsen, 1998). Additionally, listwise deletion can bias parameter estimates if data is missing at random (MAR) or missing not at random (MNAR) (Roth, 1994).

3.2 Mean Imputation

This is the easiest way to impute which means replacing each missing value with the mean of the observed values for that variable. This method simply imputes the mean of the observed data. Mean imputation is known to be a bad strategy, and the user should be aware of the implications (Buuren & Groothuis-Oudshoorn, 2011).

3.3 Multiple Imputation

Multiple imputation is a useful technique when dealing with data sets with missing values. It is a popular way to handle missing data under MAR assumption (Little and Rubin, 2002). Instead of filling in a single value for each missing value, using Rubin's (1987) multiple imputation method each missing value is replaced with a set of plausible values representing the uncertainty about the right value to impute (Yuan, 2010). The precision of the study associations is commonly overestimated with single imputation due to obtaining very low estimates of the standard error, while multiple imputation results in correct estimates of the standard errors (Koopman et al., 2008).

In multiple imputation, the missing data are stochastically imputed m times. In the commonest approach, the m completed data sets are then analyzed using methods appropriate for complete data, and using Rubin's rule, the m results are combined (Rubin, 1987). There seems to be a general consensus that more modern approaches such as multiple imputation or full information maximum likelihood are preferable to traditional approaches such as listwise deletion (Buhi & Goodson, 2008).

4. Example: Mental Health facility

Data are presented from a mental health facility in North Carolina that works as a healing community to help individuals with a mental health challenge or emotional distress to learn new ways to gain coping skills, learn to become independent, and attain fulfillment in life through a comprehensive program. Data analysis was conducted using SAS 9.4 and R 3.2.0.

This research tries to predict the length of stay of the residents of this facility. This study describes length of stay and other baseline variables, selects relevant variables, and selects logit functions to be used within ordinal multinomial logistic regression in the presence of missing observations.

One of the challenges of working with this dataset is that the LOS is right-skewed and truncated at zero. There also are some limits of possible LOS values due to their nature. Another challenge is having to deal with numerous baseline measures available to be used in LOS prediction. Replicated observations is another challenge as it violates the assumption of the observations independence; therefore, some problems such as overestimation of the statistical significance and underestimation of variance may arise if the correlated observations are ignored (Williams, 1995). The correlated measurements add a complexity to the statistical model which requires some adjustments. Due to the fact that those more complex models were not the purpose of this study, for this analysis aggregation is used to take care of the correlated observations issues.

The original dataset included 322 observations of the 40 baseline variables. In order to start this analysis, the baseline measures that effectively predict length of stay needed to be identified to be used in the future models. For this initial analysis, a log-linear Poisson regression model was applied to account for the right-skewness inherent in length of stay, and LASSO estimation was used to account for multicollinearity and to select the appropriate predictors using SAS. Missing values were eliminated using listwise deletion at this stage for simplicity so this part of analysis was done on 242 complete observations.

It was found that seven variables can be used to effectively predict length of stay with the correlation of about 0.806 between predictors and observed lengths of stay. These variables are health survey (HS), spirituality survey (SS) representing the measure of spirituality, race / ethnicity indicators for Caucasian (C) and Hispanic (H), marital status (MS), depression (D), and anxiety (A). The formula for predicting length of stay is

$$\widehat{LOS} = \exp(5.10135 + 0.06884 \times HS - 0.01032 \times SS - 0.11066 \times C + 0.54479 \times H + 0.08172 \times MS + 0.00842 \times D + 0.00101 \times A)$$

Eight additional variables were selected with lighter restrictions giving the correlation of about 0.843 between predictions and observed lengths of stay. They are the indicators for Schizophrenia, Personality Disorder, Future Scale: Snyder Hope Scale (Futs_00), Mental Health Recovery Measure at baseline (mhrm_00), Obsessive-Compulsive Disorder Indicator at baseline (O_C0), Health Outcomes Survey at baseline (HOS0), Global Scale Inventory at baseline (GSI0), and Positive Symptom Distress Index measured at baseline (PSDI0).

4.1 Model Missingness

This part of study was done to figure out how serious the missing data are using baseline measures. Also to find out whether data are missing at random or not, using logistic regression and if they are not missing at random, to find the commonality amongst observations with a specific value for one variable having missing values for the other variables.

Answering the question of whether the values that are missing are associated with any characteristics of the clients who have missing values or not via this model will justify whether more appropriate methods for accounting for missing data should be applied or not. Applying logistic regression to model probability of missingness using all baseline predictors will answer the question about the nature of the missing data.

Logistic regression is useful for predicting the presence or absence of a characteristic or an outcome based on values of a set of predictor variables. Through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or discrete, or any combination of both types and they do not necessarily have normal distributions (Hosmer & Lemeshow, 2013). The most common form of logistic regression which is used for this study uses the logit link function which gives us the logistic regression equation as

$$logit\ (p_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}.$$

After applying the logistic regression to the baseline measures, the only variable that has minimal association with missingness is gender telling us that men were marginally more likely to have missing values.

None of the other variables were significant that tells us that the data might be at least MAR but we cannot justify Missing Completely At Random (MCAR). Therefore, applying listwise deletion is appropriate here, but applying other techniques of handling missing data which were discussed above, especially multiple imputation, is interesting and might provide more information.

Using different software packages, SAS and R, and reporting different options researchers currently have under each of them are informative for both this research and for other researchers' use.

Within SAS, listwise deletion is default for most of the procedures including the ones used for this analysis within ordinal multinomial logistic models so no specific procedure needs to be added for listwise deletion. Mean imputation is easily done using PROC STANDARD. On the other hand, multiple imputation is more complicated and needs be done in three steps; first using PROC MI to impute data, then running the actual analysis (i.e., PROC LOGISTIC for this data analysis), and finally PROC MIANALYZE to pool the results from all imputations together and get the final results. Unfortunately PROC MI/PROC MIANALYZE is not compatible with ordinal models and it gives only one intercept and uses t-test when trying to test the significance of the independent variables while under listwise deletion and mean imputation, there are multiple intercepts as there should be within these ordinal models and the Z-tests and Wald tests are used to the test of significance of the model and the parameters.

Due to all these limitations for the multiple imputation in SAS, the same analyses have been done in R which is more straightforward. As in SAS, listwise deletion is default in R for this type of analysis. Mean Imputation is easily done using MICE package specifying "MEAN" option or easily by writing a function replacing missing observations with the mean of other observations within each variable. Multiple imputation can be easily done using MICE package specifying "NORM" option for the "Bayesian Linear Regression" type of imputation.

There are other options for multiple imputation within the MICE package, but the NORM option is being used to be comparable to SAS results mentioned above which uses Markov chain Monte Carlo (MCMC) method in multiple imputation procedure. Other imputation options within MICE package in R are predictive mean matching (PMM), non-Bayesian linear regression (NORM.NOB), Two-level linear model (2L.NORM), logistic regression (LOGREG), polytomous (unordered) regression (POLYREG), linear discriminant analysis (LDA), and Random sample from the observed data (SAMPLE) (Buuren & Groothuis-Oudshoorn, 2011).

Although the multiple imputation method within R is also not compatible with ordinal models and it gives only one intercept and uses t-test, due to the type of output that can be extracted from this R procedure, Z-test results can be reported , same as listwise and mean imputation results, using Rubin's rule. Individual Z-values from each of the Z-tests within each imputation and the final pooled Z-values can be found in Appendix Table 1 for the cumulative logit model. Due to lack of space, the pooled Z-values using Rubin's rule for the two other logit functions are not included in the appendix.

4.2 Model Grouped Length of Stay

Instead of predicting the actual length of stay (in days) which was done as the initial analysis and involves a lot of random noise, it may be of interest to model length of stay in greater groups. For example, models can be constructed to predict the chance of a client staying for less than three months, between three and six months, etc. This "coarser" view may give a more reliable and useful indication of how long clients tend to stay based on initial measures, and it also is more interesting to the mental healing facility in different aspects including the financial planning.

The LOS was categorized into four groups which are up to three months (group 1), three to six months (group 2), six to twelve months (group 3), and finally more than twelve months (group 4). Using an Ordinal Multinomial Logistic Regression with 14 predictors, which were of the mental facility's interest in the second phase of the study, the chance of falling into each of these groups is predicted.

Test of global null hypothesis for the ordinal multinomial logistic regression model for this model is significant. The significant predictors when using listwise deletion as the technique of handling missing data under cumulative logit model were baseline Health Survey Measure on Admission (hsur_00), Depression on Admission (DEP0), Positive Symptom Distress Index (PSDI0), and the baseline measure of Obsessive Compulsive disorder (O_C0). Higher values of all of these variables result in higher chance of longer stay based on the output provided in Appendix Table 2 which shows the output from SAS. R provided the same results under the same model which its output is not included in this paper due to lack of space.

Different logit functions within these models are used in both SAS and R. In SAS, PROC GENMOD or PROC LOGISTIC can be used to perform an ordinal multinomial logistic regression model using a cumulative logit. The CLOGIT option (LINK=CLOGIT) needs to be added.

PROC NLMIXED or PROC CATMOD can be used to perform the adjacent categories logit model, but due to the fact that there still is not a built in procedure in SAS for this type of analysis, the likelihood functions need to be typed within the NLMIXED procedure which can be time consuming specially when there are a lot of independent variables in the model. PROC CATMOD is not recommended to be used for such models due to some issues it has due to being outdated.

There exist even more problems when running models using a continuation ratio logit model using PROC CATMOD which is suggested to be used by some references due to the same reason of being outdated and the issues it has which never got fixed by the SAS institute, so it does not provide very reasonable output. Agresti (2013) suggests using PROC GENMOD for the continuation-ratio logit models which performs better than PROC CATMOD. Another option when running the continuation ratio model is within PROC LOGISTIC in which various sources (e.g., Allison, 2012) demonstrate how to restructure the original dataset. With the restructured dataset and the created binary response variable PROC LOGISTIC produces the same results as NLMIXED. Within this procedure the PARAM=GLM coding in the CLASS statement should be used rather than as an option on the MODEL statement (High, 2013).

Unfortunately most of these options for the adjacent-category and continuation ratio logit models are either time consuming to perform in SAS or not compatible with some of the missing data handling techniques such as multiple imputation.

Within R, performing ordinal multinomial logistic models using different logit functions is easier. They can be done using the package "VGLM". The option FAMILY=CUMULATIVE needs to be added for cumulative logit functions, FAMILY=ACAT should be added for the adjacent categories logit models, and finally FAMILY=CRATIO or FAMILY=SRATIO should be added for the continuation ratio logit models when applying listwise deletion and mean imputation as the missing data handling techniques. FAMILY=SRATIO is not compatible with multiple imputation so in this case the option that can be used is FAMILY=CRATIO.

The outputs from mean imputation within different logit models are not included here as it is not the preferred method of handling missing data due to many reasons such as distorting the distribution for the variables with missing observations, biasing the standard errors, and not preserving the relationships among variables. The results in terms of significance of predictors were also different from the other methods and it resulted in getting more significant predictors.

The results from R using multiple imputation for all three logit functions, Cumulative logit, Adjacent-Categories logit, and Continuation-Ratio logit, are shown in Appendix Tables 3, 4, and 5 respectively.

Appendix Table 6 summarizes all of the available procedures within SAS and R for the combinations of different logit functions and missing data handling techniques.

5. Conclusion

In this paper a comparison between different logit functions within an ordinal multinomial logistic regression in SAS and R is presented along with the comparison of three missing data handling techniques.

Due to the real situations dealt with when analyzing the mental healing facility data and based on the lack of studies using ordinal multinomial logistic models for modeling the LOS, this study focused on this comparison and providing information in terms of available procedures in two different software packages for researchers willing to model ordered categorical response variables in presence of missing data. Unfortunately use of ordered information is not very common among researchers in different fields probably due to these models' complexity and the software packages' limitations, but this should not stop using such models when dealing with ordered responses due to the higher power of the ordinal hypothesis tests and the possibility of taking into consideration the ordinal nature of the response when interpreting the results. The more research done and published using these models provide more resources for researchers to use them which was one of the reasons of writing this paper.

Different options within SAS and R are provided in this study and using them on the real data shows that using listwise deletion along with the cumulative logit is the easiest one to perform both in SAS and R. However, sometimes due to the type of results we are hoping to get and also based on a specific interpretation we want to present, other logit functions might be more appropriate. Also, when there are some characteristics of some variables involved in the type of missingness, applying more advanced missing data handling techniques specifically multiple imputation is recommended. Using MICE package in R is highly recommended for performing multiple imputation due to its simplicity and also giving the z-test results from each imputation that can be pooled together using Rubin's rule. For SAS users, PROC MI/PROC MIANALYZE can be used for this type of imputation.

In this study the mean imputation missing data handling technique is also considered which is not recommended in general based on different studies and the results author observed from its output not aligning with other missing techniques leading to biased parameter estimates.

Applying multiple imputation within some of the logit functions made the analysis face some issues, but some of them were fixable by mixing some of the procedures together within SAS and R which are all reported in this study (See table 6). For adjacent-categories and continuation ratio logit models, using VGLM package is recommended over SAS procedures.

All in all, more code needs to be written and more procedures need to be built into different software packages, especially SAS, due to the limitations this author faced using this software package when mixing some of the logit functions with multiple imputation. Having more options in terms of statistical software will enable researchers using these ordinal models along with the appropriate technique for handling missing data to increase the power of studies involving ordered data and missing observations.

References

Abreu, M. N. S., Siqueira, A. L., & Caiaffa, W. T. (2009). Ordinal logistic regression in epidemiological studies. *Revista de Saude publica*, *43*(1), 183-194.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). New York: Wiley.

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). New York: Willey.

Allison, P. D. (2012). *Logistic regression using SAS: Theory and application*. SAS Institute.

Buhi, E. R., Goodson, P., & Neilands, T. B. (2008). Out of sight, not out of mind: strategies for handling missing data. *American journal of health behavior*, *32*(1), 83-92.

Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, *45*(3).

Cheema, J. R. (2014). A Review of Missing Data Handling Methods in Education Research. *Review of Educational Research*, *84*(4), 487-508.

Gibson, N. M., & Olejnik, S. (2003). Treatment of missing data at the second level of hierarchical linear models. *Educational and Psychological Measurement*, *63*(2), 204-238.

High, R. Models for Ordinal Response Data (2013). *SAS Global Forum, Paper 445-2013*.

Hosmer Jr, D. W., & Lemeshow, S. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons.

Koopman, L., van der Heijden, G. J., Grobbee, D. E., & Rovers, M. M. (2008). Comparison of methods of handling missing data in individual patient data meta-analyses: an empirical example on antibiotics in children with acute otitis media. *American journal of epidemiology*, *167*(5), 540-545.

Lall, R., Campbell, M. J., Walters, S. J., Morgan, K., & Co-operative, M. C. (2002). A review of ordinal regression models applied on health-related quality of life assessments. *Statistical methods in medical research*, *11*(1), 49-67.

Lee, K. J., Galati, J. C., Simpson, J. A., & Carlin, J. B. (2012). Comparison of methods for imputing ordinal data using multivariate normal imputation: a case study of non-linear effects in a large cohort study. *Statistics in medicine*, *31*(30), 4164-4174.

Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis With Missing Data (2nd ed.)*, New York: Wiley.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel psychology*, *47*(3), 537-560.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate behavioral research*, *33*(4), 545-571.

Williams, R. L. (1995). Product-limit survival functions with correlated survival times. *Lifetime data analysis*, *1*(2), 171-186.

Yuan, Y. C. (2010). Multiple imputation for missing data: Concepts and new development (Version 9.0). *SAS Institute Inc, Rockville, MD*, *49*.

Appendix

Table 1

Pooled Z-values from 10 imputations using Rubin's rule

Parameter	Imp1 zvalue	Imp2 zvalue	Imp3 zvalue	Imp4 zvalue	Imp5 zvalue	Imp6 zvalue	Imp7 zvalue	Imp8 zvalue	Imp9 zvalue	Imp10 zvalue	Pooled zvalue
(Intercept):1	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
(Intercept):2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
(Intercept):3	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
SA	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63
Personality	2.18	2.18	2.18	2.18	2.18	2.18	2.18	2.18	2.18	2.18	2.18
futs_00	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89
hsur_00	3.34	3.34	3.34	3.34	3.34	3.34	3.34	3.34	3.34	3.34	3.34
mhrm_00	2.01	2.01	2.01	2.01	2.01	2.01	2.01	2.01	2.01	2.01	2.01
Sibr_00	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
race2	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39
race3	2.84	2.84	2.84	2.84	2.84	2.84	2.84	2.84	2.84	2.84	2.84
race4	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63
race5	2.22	2.22	2.22	2.22	2.22	2.22	2.22	2.22	2.22	2.22	2.22
race6	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
race7	2.65	2.65	2.65	2.65	2.65	2.65	2.65	2.65	2.65	2.65	2.65
marital_status2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
marital_status3	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
marital_status4	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
O_C0	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13
DEP0	3.37	3.37	3.37	3.37	3.37	3.37	3.37	3.37	3.37	3.37	3.37
ANX0	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19
HOS0	6.89	6.89	6.89	6.89	6.89	6.89	6.89	6.89	6.89	6.89	6.89
GSI0	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63
PSDI0	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07

Table 2

Ordinal Multinomial Logistic Regression - Cumulative Logit (Listwise Deletion)

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	4	1	-4.1602	1.1393	13.3339	0.0003
Intercept	3	1	-2.7307	1.1142	6.0066	0.0143
Intercept	2	1	-1.0112	1.1002	0.8448	0.3580
SA	0	1	0.3520	0.2740	1.6503	0.1989
Personality	0	1	-0.5296	0.2961	3.1999	0.0736
futs_00		1	0.1486	0.3530	0.1771	0.6739
hsur_00		1	0.8512	0.2584	10.8472	0.0010
mhrm_00		1	-0.1174	0.3099	0.1435	0.7048
Sibr_00		1	0.0699	0.1610	0.1885	0.6641
race	Asian/Paci	1	1.4664	1.3156	1.2424	0.2650
race	Hispanic	1	1.8181	1.1461	2.5164	0.1127
race	Middle Eas	1	-12.4552	1027.2	0.0001	0.9903
race	Multi-raci	1	-10.4533	1027.2	0.0001	0.9919
race	Native Ame	1	-0.8489	1.0789	0.6191	0.4314
marital_status	D	1	0.8101	1.3384	0.3663	0.5450
marital_status	M	1	-0.5878	0.5349	1.2075	0.2718
O_C0		1	0.1049	0.0389	7.2859	0.0069
DEP0		1	0.1196	0.0400	8.9453	0.0028
ANX0		1	0.0318	0.0394	0.6490	0.4205
HOS0		1	0.0517	0.0483	1.1475	0.2841
GSI0		1	-0.1381	0.1123	1.5100	0.2191
PSDI0		1	-1.0958	0.2854	14.7358	0.0001

Table 3

Ordinal Multinomial Logistic Regression - Cumulative Logit (Multiple Imputation)

Parameter	est	se	t	df	Pr(>jtj)	lo 95	hi 95	nmis	fmi	lambda
(Intercept):1	11.92	1015.17	0.01	864.92	0.99	-1980.56	2004.41		0	0
(Intercept):2	13.41	1015.17	0.01	864.92	0.99	-1979.07	2005.89		0	0
(Intercept):3	14.71	1015.17	0.01	864.92	0.99	-1977.77	2007.19		0	0
SA	0.4	0.24	1.68	731.96	0.09	-0.07	0.87	0	0.04	0.04
Personality	-0.51	0.29	-1.77	245.12	0.08	-1.07	0.06	0	0.16	0.16
futs_00	-0.33	0.35	-0.96	112.64	0.34	-1.02	0.35	1	0.27	0.26
hsur_00	-0.73	0.25	-2.94	145.88	0	-1.22	-0.24	5	0.23	0.22
mhrm_00	0.37	0.3	1.23	228.78	0.22	-0.22	0.95	3	0.17	0.16
Sibr_00	-0.05	0.16	-0.29	79.06	0.77	-0.37	0.28	4	0.33	0.31
race2	1.38	2.49	0.55	51.52	0.58	-3.61	6.37		0.42	0.4
race3	3.13	2.1	1.49	38.06	0.15	-1.13	7.38		0.49	0.47
race4	1.8	2.3	0.78	55.96	0.44	-2.81	6.42		0.4	0.38
race5	4.1	2.78	1.48	36.71	0.15	-1.53	9.73		0.5	0.47
race6	15.95	1015.17	0.02	864.92	0.99	-1976.53	2008.43		0	0
race7	3.04	2.41	1.26	38.05	0.22	-1.84	7.92		0.49	0.47
marital_status2	-14.9	1015.17	-0.01	864.92	0.99	-2007.38	1977.57		0	0
marital_status3	-13.85	1015.17	-0.01	864.92	0.99	-2006.33	1978.62		0	0
marital_status4	-14.19	1015.16	-0.01	864.92	0.99	-2006.66	1978.29		0	0
O_C0	-0.06	0.08	-0.8	10.44	0.44	-0.23	0.11	62	0.9	0.88
DEP0	-0.08	0.09	-0.89	9.77	0.39	-0.29	0.12	62	0.92	0.9
ANX0	-0.02	0.07	-0.3	11.74	0.77	-0.17	0.13	62	0.86	0.84
HOS0	-0.05	0.06	-0.75	17.08	0.46	-0.18	0.08	62	0.73	0.7
GSI0	0.15	0.2	0.76	13.07	0.46	-0.28	0.58	62	0.82	0.8
PSDI0	0.54	0.42	1.28	12.37	0.22	-0.38	1.46	62	0.84	0.82

Table 4

Ordinal Multinomial Logistic Regression - Adjacent Categories Logit (Multiple Imputation)

Parameter	est	se	t	df	Pr(>jtj)	lo 95	hi 95	nmis	fmi	lambda
(Intercept):1	13.2	994.39	0.01	864.92	0.99	-1938.5	1964.89		0	0
(Intercept):2	13.53	994.39	0.01	864.92	0.99	-1938.16	1965.22		0	0
(Intercept):3	13.77	994.39	0.01	864.92	0.99	-1937.92	1965.46		0	0
SA	0.28	0.14	2	691.53	0.05	0.01	0.55	0	0.05	0.05
Personality	-0.28	0.16	-1.73	262.64	0.09	-0.61	0.04	0	0.15	0.15
futs_00	-0.2	0.19	-1.03	175.6	0.31	-0.58	0.18	1	0.2	0.2
hsur_00	-0.43	0.15	-2.94	148.28	0	-0.72	-0.14	5	0.23	0.22
mhrm_00	0.24	0.17	1.38	238.46	0.17	-0.1	0.58	3	0.17	0.16
Sibr_00	-0.03	0.09	-0.28	104.23	0.78	-0.21	0.16	4	0.28	0.27
race2	0.76	1.3	0.58	80.91	0.56	-1.83	3.35		0.33	0.31
race3	1.62	1.12	1.44	53.8	0.16	-0.63	3.88		0.41	0.39
race4	0.74	1.29	0.57	63.38	0.57	-1.84	3.31		0.37	0.35
race5	2.16	1.52	1.42	49.85	0.16	-0.89	5.2		0.43	0.4
race6	14.82	994.39	0.01	864.92	0.99	-1936.88	1966.51		0	0
race7	1.54	1.3	1.18	49.81	0.24	-1.08	4.15		0.43	0.4
marital_status2	-14.38	994.39	-0.01	864.92	0.99	-1966.07	1937.31		0	0
marital_status3	-13.83	994.39	-0.01	864.92	0.99	-1965.52	1937.86		0	0
marital_status4	-13.96	994.39	-0.01	864.92	0.99	-1965.65	1937.73		0	0
O_C0	-0.03	0.04	-0.8	10.93	0.44	-0.12	0.06	62	0.88	0.86
DEP0	-0.04	0.05	-0.89	10.2	0.39	-0.15	0.07	62	0.91	0.89
ANX0	-0.01	0.04	-0.3	12.23	0.77	-0.09	0.07	62	0.85	0.82
HOS0	-0.03	0.03	-0.78	18.55	0.45	-0.1	0.04	62	0.7	0.67
GSI0	0.08	0.11	0.7	13.76	0.5	-0.16	0.31	62	0.81	0.78
PSDI0	0.32	0.25	1.25	12.47	0.23	-0.23	0.86	62	0.84	0.82

Table 5

Ordinal Multinomial Logistic Regression - Continuation Ratio Logit (Multiple Imputation)

Parameter	est	se	t	df	Pr(>jtj)	lo 95	hi 95	nmis	fmi	lambda
(Intercept):1	-12.46	973.4	-0.01	864.92	0.99	-1922.96	1898.04		0	0
(Intercept):2	-13.17	973.4	-0.01	864.92	0.99	-1923.67	1897.34		0	0
(Intercept):3	-13.85	973.4	-0.01	864.92	0.99	-1924.36	1896.65		0	0
SA	-0.44	0.2	-2.21	770.94	0.03	-0.83	-0.05	0	0.03	0.03
Personality	0.4	0.24	1.67	277.41	0.1	-0.07	0.86	0	0.15	0.14
futs_00	0.35	0.28	1.22	131.44	0.22	-0.22	0.91	1	0.25	0.23
hsur_00	0.65	0.21	3.02	101.22	0	0.22	1.07	5	0.29	0.27
mhrm_00	-0.35	0.25	-1.41	233.7	0.16	-0.84	0.14	3	0.17	0.16
Sibr_00	0	0.13	0.01	81.1	1	-0.27	0.27	4	0.33	0.31
race2	-0.64	1.94	-0.33	94.52	0.74	-4.49	3.21		0.3	0.28
race3	-2.32	1.63	-1.43	58.62	0.16	-5.58	0.94		0.39	0.37
race4	-1.23	1.82	-0.67	83.63	0.5	-4.85	2.4		0.32	0.3
race5	-3.28	2.16	-1.52	55.41	0.13	-7.61	1.04		0.4	0.38
race6	-15.59	973.4	-0.02	864.92	0.99	-1926.09	1894.91		0	0
race7	-2.14	1.89	-1.13	52.81	0.26	-5.94	1.65		0.41	0.39
marital_status2	14.8	973.4	0.02	864.92	0.99	-1895.7	1925.3		0	0
marital_status3	14.09	973.4	0.01	864.92	0.99	-1896.41	1924.59		0	0
marital_status4	14.29	973.4	0.01	864.92	0.99	-1896.21	1924.79		0	0
O_C0	0.04	0.06	0.77	11.12	0.46	-0.08	0.17	62	0.88	0.86
DEP0	0.06	0.07	0.88	10.38	0.4	-0.09	0.21	62	0.9	0.88
ANX0	0.02	0.05	0.39	12.74	0.7	-0.09	0.14	62	0.83	0.81
HOS0	0.03	0.05	0.73	19.49	0.48	-0.07	0.13	62	0.69	0.66
GSI0	-0.11	0.15	-0.72	14.26	0.48	-0.43	0.21	62	0.79	0.77
PSDI0	-0.45	0.36	-1.26	12.29	0.23	-1.22	0.33	62	0.84	0.82

Table 6

SAS and R procedures for Ordinal Multinomial Logistic models with different missing data handling techniques

	Missing Data Handling Technique		
	Listwise Deletion	Mean Imputation	Multiple Imputation
SAS			
Cumulative Logit	proc genmod/logistic link=clogit	proc genmod/logistic & proc standard	proc genmod/logistic & proc mi/mianalyze
Adjacent Category	proc nlmixed (enter likelihood) or catmod	proc nlmixed(enter likelihood) or catmod & proc standard	proc nlmixed(enter likelihood) or catmod & proc mi/mianalyze
Continuation Ratio	proc catmod or genmod or logistic	proc catmod or genmod or logistic & proc standard	proc catmod or genmod or logistic & proc mi/mianalyze
R			
Cumulative Logit	vglm (family=cumulative)	vglm (family=cumulative) (mice, mean) or write function	vglm (family=cumulative) mice, norm
Adjacent Category	vglm (family=acat)	vglm (family=acat) (mice, mean) or write function	vglm (family=acat) mice, norm
Continuation Ratio	vglm (family=cratio/sratio)	vglm (family=cratio/sratio) (mice, mean) or write function	vglm (family=cratio) mice, norm