

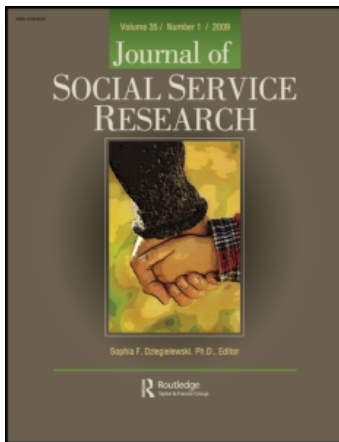
This article was downloaded by: [Petrucci, Carrie J.]

On: 5 March 2009

Access details: Access Details: [subscription number 909201879]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Social Service Research

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t792306968>

A Primer for Social Worker Researchers on How to Conduct a Multinomial Logistic Regression

Carrie J. Petrucci ^{ab}

^a EMT Associates, Inc., Encino, CA ^b School of Social Work, California State University, Los Angeles

Online Publication Date: 01 April 2009

To cite this Article Petrucci, Carrie J.(2009)'A Primer for Social Worker Researchers on How to Conduct a Multinomial Logistic Regression',Journal of Social Service Research,35:2,193 — 205

To link to this Article: DOI: 10.1080/01488370802678983

URL: <http://dx.doi.org/10.1080/01488370802678983>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A Primer for Social Worker Researchers on How to Conduct a Multinomial Logistic Regression

Carrie J. Petrucci

ABSTRACT. The purpose of this article is to explain how to conduct a multinomial logistic regression (MLR) to increase its usage among social work researchers. A challenge for social work researchers carrying out practice-based research is to incorporate statistical analyses that are comprehensible to clinicians. Another challenge is that categorical variables, as opposed to continuous, commonly occur in clinical settings, yet their usage is seldom taught in social work education. This article will discuss MLR, a categorical data analysis used when there are three or more unordered categories in the outcome variable. Multinomial logistic regression determines differences across client groups, and can be useful in assessment, case planning, and examination of outcomes.

KEYWORDS. Evidence-based practice, multinomial logistic regression (MLR), practice-based research, statistical techniques.

One of the major challenges for social work researchers is to incorporate statistical analyses into practice-based research that are both comprehensible and useful to social work clinicians and practitioners (Gingerich, 1990; Goldstein, 2007; Kazi, 2003; Kirk, 1990; Reid & Fortune, 1992; Simpson, Williams, & Segall, 2007; Thomas & Rothman, 1990; Thyer, 2007; Wood, 1990). This means that the findings must be applicable to practice with a minimal number of limitations and caveats. For example, a natural logarithm transformation of a variable in a statistical analysis necessitates further explanation and unspecified limitations if it is to be applied in agency practice. If social work researchers

want research findings utilized in real-world settings, the findings must be based on an analysis that is discernible to a wide range of practitioners, regardless of their background in statistics. This article is an effort to broaden social work researchers' skills in a statistical technique that may be quite useful to practitioners in examining and improving clinical interventions.

An additional challenge to practice-based research is that categorical variables, as opposed to continuous or linear variables, commonly occur in social science settings (Long, 1997), yet their usage is seldom taught in required research methods and statistics classes in graduate social work education. A common example

Carrie J. Petrucci, PhD, is a senior research associate at EMT Associates, Inc., Encino, CA, and a part-time lecturer and field instructor at the School of Social Work, California State University, Los Angeles.

Address correspondence to Carrie J. Petrucci: ubiquitus@aol.com or cpetrucci@emt.org.

Portions of this paper were presented at the American Evaluation Association Annual Conference, November 2006, and at the 11th Annual Society for Social Work Researchers Conference in January 2007. The author wishes to thank Mansoor Kazi, Thomas Nochajski, Moisés Próspero, Audrey Hickert and three anonymous reviewers for their suggestions on earlier drafts of this article.

of a categorical variable is a two-level variable that includes those who complete a program and those who terminate early. A commonly used continuous or linear variable might be the change score on a standardized assessment or scale. Social work students are typically exposed to group comparison statistics such as *t* tests and analysis of variance (ANOVA), which assume a continuous or numerical normally distributed outcome variable. Social work curriculums seldom spend any time on categorical analyses beyond chi-square statistical testing. This leaves social work researchers unprepared to answer common questions that may be of great interest in practice. Examples of possible questions include: what are the characteristics that differentiate clients who successfully complete a program from those who fail to complete a program, and what characteristics are common to both? These characteristics could then be shared with practitioners to document successes for specific subgroups of clients as well as determine if program or implementation improvements are needed for subgroups of clients not doing well. Logistic regression could be used for this analysis. While logistic regression is similar to linear regression, it requires exposure to concepts common to categorical data analyses that many social work researchers and practitioners are not getting in their graduate education.

This article will discuss one type of categorical data analysis, multinomial logistic regression (MLR). Logistic regression is appropriate when there are two mutually exclusive categories in the dependent or outcome variable (program success/program failure). Multinomial logistic regression is used when there are three or more categories to the dependent or outcome variable (Hosmer & Lemeshow, 2000). These categories can be ordered (referred to as ordinal regression models and not covered in this article) or unordered (referred to as nominal, polychotomous, polytomous, discrete choice, or multinomial models, and the topic of this article) (Hosmer & Lemeshow, 2000; Long, 1997).

An MLR can be used to analyze predictors for an unordered group classification, such as children who experienced no abuse, physical abuse only, or sexual abuse only (Mandell, Walrath, Manteuffel, Sgro, & Pinto-Martin, 2005). While

the no abuse category is certainly less severe than the two abuse categories, there is no inherent order assumed to the severity of the physical or sexual abuse categories, and the no abuse category provides a basis for comparison to the other two groups. Multinomial logistic regression is similar to multiway contingency table and log-linear analyses, but is more intuitive to interpret, particularly when there are several independent variables being examined with a dependent variable (Tabatchnick & Fidell, 2007).

An advantage of MLR (and logistic regression) is its use of odds ratios as estimators for the predictor variables. This provides researchers and practitioners with a more intuitive interpretation to the final model produced. Another advantage to MLR is that both categorical and continuous independent variables can be incorporated as predictors. This is not the case in discriminant analysis, which also classifies group membership based on a combination of variables (Stevens, 1996). In addition, discriminant analysis requires normally distributed independent variables, and is generally less preferred over logistic regression models (Howell, 2002).

Few studies have examined differences in performance for multinomial and ordinal regression models compared to linear regression or discriminant analysis (Hossain, Wright, & Petersen, 2002). Different research questions call for different analyses and the required assumptions of the data also influence the most appropriate analyses. When the normality assumption for the dependent variable cannot reasonably be assumed, and the outcome is three or more unordered categories, MLR may be the most appropriate choice of analysis. Multinomial logistic regression can also be extended into more sophisticated statistical analyses that incorporate time as a factor, such as a competing risks-model and survival analysis (Allison, 1995). These types of analyses examine time to an event to determine the risk for multiple competing outcomes, while also modeling predictors for those outcomes.

The purpose of this article is to present a step-by-step example of how to conduct an MLR using SPSS 14.0 and to familiarize social work researchers with the basics of MLR in order to increase its appropriate usage as a viable tool in practice-based research settings. A brief review

TABLE 1. Number of Articles Found on Multinomial Logistic Regression (MLR), Logistic Regression, and Regression in Selected Databases in January 2008

Database	MLR	Logistic Regression	Regression
Social Work Abstracts	21	344	1,149
Social Services Abstracts	70	901	1,574
Sociological Abstracts	256	1,505	5,848
PsychInfo	666	8,361	29,858
Medline	1,229	56,141	133,111

of studies in the social work and social services literature that have used MLR will be presented, followed by a step-by-step MLR example.

EXAMPLES OF MULTINOMIAL LOGISTIC REGRESSION IN THE SOCIAL WORK LITERATURE

Table 1 presents the results of database searches conducted in January 2008 to determine how common MLR is across various helping professions. Searches on each type of analyses in the abstracts were performed (“multinomial” OR “multinomial logit” OR “multinomial logistic regression,” “logistic regression NOT multinomial,” and “regression NOT multinomial, NOT logistic”). As is immediately evident, the greatest number of MLR analyses has been published in Medline, followed by PsychInfo. Health researchers have utilized categorical analyses for many years and it is included in basic statistics courses (see Dunn & Clark, 2001). Social work and social services publications have rarely published MLR analyses, although they are definitely on the rise (since 2001, 19 out of 21 of the articles in Social Work Abstracts and 58 out of 70 of the articles in Social Services Abstracts occurred). There are clearly more articles that use logistic regression and regression in social work and social services, speaking to the greater familiarity with logistic and linear regression. The greater usage of MLR among health, sociology, and psychology researchers argues for its potential usefulness in social work research.

MULTINOMIAL LOGISTIC REGRESSION EXAMPLE

Data from an example from a statewide evaluation of 45 child-abuse treatment agencies conducted by the author will be used to explain how an MLR is performed and interpreted. This real-world data was used as an example to illustrate some of the common strengths and limitations that can be encountered in an MLR analysis. SPSS 14.0 was used for the example. This is intended as a “how-to” demonstration for those with little to no familiarity with interpreting categorical data analyses. Familiarity with how to interpret regression coefficients and odds ratios will be helpful and is reviewed briefly next.

Regression Coefficients

Regression analyses determine the best set of predictors or independent variables for a numerical or scaled dependent or outcome variable, such as the score on a standardized assessment. Regression is often referred to as “fitting the best line” to the data. In linear regression, the predictors are computed based on the linear relationship between each predictor and outcome variable. In other words, lower amounts of a predictor variable should result in lower amounts of an outcome, and higher amounts of a predictor variable should result in a higher amount of the outcome. The relationship between the predictor and outcome variables produces a line that goes either up or down. Regression coefficients, often symbolized as b (unstandardized) or β (standardized and referred to as *beta*), are the slope of the line, or the amount of change in the dependent variable (symbolized as Y) based on a one-unit change in the predictor or independent variable (symbolized as X) (Abu-Bader, 2006; Dunn & Clark, 2001; Howell, 2002). However, when the independent variables are categorical, regression coefficients become difficult to interpret. Odds ratios are a very useful alternative (Dunn & Clark, 2001; Howell, 2002; Rosenthal, 2001).

Odds and Odds Ratios

Odds ratios can be used when the dependent or outcome variable has two categories, such as

TABLE 2. Odds Ratio Table Depicting Success and Failure Across a Treatment and Comparison Group

Outcome (Event)	Treatment (Group)	Comparison (Group)	Total
Success	115 = a	60 = b	175 = $a + b$
Failure	20 = c	40 = d	60 = $c + d$
Total	135 = n_1	100 = n_2	235 = n

success or failure, and there are two or more groups of clients. Odds ratios are easily computed from the beta coefficients in logistic regression and MLR using the exponent function on any calculator (Hosmer & Lemeshow, 2000). When interpreting odds ratios, it is important to differentiate between the *event* or outcome variable (success/failure), and the independent or group variable (treatment/comparison groups). See Table 2 for an example.

The odds for a particular event, in this case treatment success, are computed across both groups as follows:

$$\begin{aligned} \text{Odds for Success} &= \frac{\text{total number who succeeded}}{\text{total sample} - \text{total succeeded}} \\ &= \frac{175}{235 - 175} = \frac{175}{60} = 2.9 \end{aligned}$$

To interpret *odds*, we are looking at the odds of the numerator (total number who succeeded) compared to the denominator (total failed). Thus, *treatment success is 2.9 times greater than treatment failure across both groups*. For just the treatment group, the odds for success are 5.7 to 1 (computed by dividing $115/(135 - 115)$ from Table 2 using the formula above). This can be interpreted as follows: For every 5.7 people who succeeded in the treatment group, 1 person failed. For the comparison group, the odds for success were 1.5 to 1 (computed by dividing $60/100 - 60$). For every 1.5 people who succeeded in the comparison group, one person failed. The *odds ratio* then is the ratio of odds between the two groups being compared on an outcome, as

follows:

Odds Ratio

$$\begin{aligned} &= \frac{a/c \text{ (treatment group odds for success)}}{(b/d \text{ comparison group odds for success})} \text{ or } \frac{ad}{bc} \\ &= \frac{115/20}{60/40} \text{ or } \frac{(115)(40)}{(60)(20)} = \frac{5.75}{1.5} \text{ or } \frac{4600}{1200} = 3.8 \end{aligned}$$

Mathematically, the odds ratio between the two groups can be computed by either dividing to get the odds for each group or by multiplying the cross products. Here we see that the odds of success for those in the treatment group compared to the comparison group was 3.8 to 1. Said another way, *people in the treatment group were 3.8 times as likely to succeed than those in the comparison group* (Dunn & Clark, 2001; Rosenthal, 2001). This will sometimes be stated as “3.8 times *more* likely to succeed” in journal articles or textbooks. Hosmer and Lemeshow (2000) point out that “more” or “less” likely would only be true when the odds ratio is believed to be similar to the relative risk. This is thought to be the case only when the outcome is a rare event, occurring for 10 percent or less of the population. However, if the odds ratio is used as a “broadstroke estimate of effect” rather than a precise estimate of the likelihood of an event, then use of the phrases “more” or “less” likely is acceptable.

Another useful aspect of odds ratios is the ability to compute the reciprocal in order to determine the odds of the opposite event, which in this case would be failure rather than success. If the odds ratio for success for the treatment group is 3.8, taking the reciprocal of 3.8 provides the odds for failure for the treatment group, which is .26 ($3.8 \times .26 = 1$). Odds ratios less than one are interpreted as *less likely* and are less easily interpreted. In this case, people in the treatment group were .26 times less likely to fail than people in the comparison group. When using the reciprocal, it is important not to confuse the event or outcome and the two groups being compared (Dunn & Clark, 2001; Rosenthal, 2001).

Research Questions

Similar to linear regression, MLR can be used for classification purposes. It is capable of

answering questions such as which risk or protective factors predict group membership in three or more unordered mutually exclusive groups. This classification information can subsequently be used in practice for assessment purposes, to determine appropriate levels of an intervention, or to link program outcomes to different groups of clients. Health researchers have used MLR extensively, largely because of its appropriateness to health issues (analyzing competing unranked causes of health issues). Similarly, in an example from mental health, researchers examined differential predictors for adolescents who were suicidal ideators, suicidal attempters, and nonsuicidal (Wild, Flisher, & Lombard, 2004).

In the example presented here, 45 child-abuse treatment agencies agreed to complete the same series of intake and follow-up assessments for all new clients over a 6-month period. One instrument to measure the child's functioning, the Children's Global Assessment Scale (CGAS) (Shaffer et al., 1983), was selected for the analysis. When looking at the pre- and posttest scores of the CGAS, three distinct groups of children emerged: children whose scores improved from intake to 6 months, those whose scores didn't change, and children whose scores got worse or deteriorated. An obvious next question became: Were there factors at intake that could be used to identify children in these three groups ahead of time, so that services could be geared proactively to those children with the greatest level of need? This became the research question for the MLR analysis presented here.

Variable Selection

Multinomial logistic regression procedures can utilize standard regression techniques to select variables (Hosmer & Lemeshow, 2000; Tabatchnick & Fidell, 2007). These include *stepwise selection*, in which statistical procedures are used to select variables that make the largest contribution to prediction of the outcome variable, or *forced entry*, in which the researcher determines which variables are included based on theory or practice. A stepwise

procedure was used in the example dataset. In MLR it is also important that no two independent variables are heavily correlated (referred to as multicollinearity). There is no definitive test for this when using categorical variables, thus conceptually similar variables were identified and one was eliminated based on which one had the smallest amount of missing data (Hosmer & Lemeshow, 2000; Tabatchnick & Fidell, 2007). Continuous variables can be included, but require checking the assumption of "linearity in the logit," or checking the relationship between the continuous variable and the outcome variable (Hosmer & Lemeshow, 2000). Interaction terms should also be explored to determine if the outcome varies by levels of an independent variable, however, this requires somewhat advanced techniques that are explained in Hosmer and Lemeshow (2000) and Tabatchnick and Fidell (2007) and are not covered here.

The variables in the final model are shown in Table 3. The dependent variable is the change in the CGAS assessed by the therapist at in-

TABLE 3. Case-Processing Summary

	N	Marginal percentage ^c
3-category change in CGAS ^a		
0 — no change	76	17.0
1 — improved	292	65.3
2 — deteriorated	79	17.7
Neglect indicated		
0 — no neglect	321	71.8
1 — neglect	126	28.2
Clinical cut-off for intake CGAS		
0 — not clinical (>61)	190	42.5
1 — clinical CGAS (<61)	257	57.5
Caregiver in extreme poverty?		
1 — Yes	84	18.8
2 — No	222	49.7
3 — Don't know	141	31.5
Valid (<i>Sample size</i>) ^b	447	100.0
Missing	0	
Total	447	
Subpopulation	12	

^aThe dependent variable is "change in CGAS." Independent variables are neglect, clinical cut-off, and poverty.

^bRatio of cases to levels of all variables = 447/10 = 45; > 20/1 preferred, suggesting adequate sample size here.

^cThese three marginal percentages squared are used in the classification formula (.17², .653², .177² = .487).

FIGURE 1. SPSS 14.0 syntax for multinomial logistic regression.

```

**Running all possible crosstabs to check for missing data and to examine all relationships**
[Notes in bold italics are author notes and are not to be included in final syntax]

CROSSTABS
  /TABLES=poverty clinicalcgas neglect [independent variables] BY cgasgroup [dependent variable]
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ CC PHI
  /CELLS=COUNT EXPECTED ROW COLUMN .
CROSSTABS
  /TABLES=poverty BY clinicalcgas neglect
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ CC PHI
  /CELLS=COUNT EXPECTED ROW COLUMN .
CROSSTABS
  /TABLES=clinicalcgas BY neglect
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ CC PHI
  /CELLS=COUNT EXPECTED ROW COLUMN .

**Selecting sample based on variable in the model with the fewest number of cases:**

USE ALL.
FILTER BY poverty .
EXECUTE .

**Syntax for the multinomial logistic regression (after variable selection process):**

LINE 1: NOMREG
LINE 2: cgasgroup [DV] (BASE=LAST ORDER=ASCENDING) BY neglect poverty clinicalcgas [IV's]
LINE 3: /CRITERIA CIN(95) DELTA(0.1) MXITER(100) MXSTEP(5) CHKSEP(20)
  LCONVERGE(0)
LINE 4: /PCONVERGE(0.000001) SINGULAR(0.00000001)
LINE 5: /MODEL
LINE 6: /INTERCEPT =INCLUDE
LINE 7: /PRINT = CLASSTABLE FIT PARAMETER SUMMARY LRT CPS MFI IC
LINE 8: /SAVE ESTPROB PREDCAT PCPROB ACPROB .

**Turning off the filter so that all cases are selected for subsequent use of the dataset**

FILTER OFF .
USE ALL .
EXECUTE .

```

take and at 6 months into treatment. The independent or predictor variables were derived from the therapists' perspective and included: neglect (yes/no); a CGAS score above or below the clinical cut-off at intake; and whether the caregiver was known to be in extreme poverty, not in extreme poverty, or at the poverty level of the caregiver was unknown by the therapist at intake (this variable was included because of its theoretical and practical importance, even though there was a significant amount of missing data).

The SPSS syntax to run an MLR is presented in Figure 1 and is included here to both assist the reader to run their own analysis as well as to enhance interpretation of the output. Tabatch-

nick and Fidell (2007) recommend running all possible comparisons (crosstabs) of the final set of variables to check for empty cells. This is where the variable selection process becomes important in balancing the number of cases lost because of missing data, and retaining variables that are theoretically important to the analysis. The "filter" command can be used after the final variables have been selected in order to eliminate all missing cases from the analysis (but not from the original dataset).

The MLR begins with Line 1: "NOMREG" is the command for a nominal or unordered MLR. In Line 2, the dependent variable ("cgasgroup") is assigned. Like logistic regression, MLR assigns a reference group to which all other levels

of the dependent variable are compared. SPSS automatically assigns the last category of the variable as the reference group. This syntax produces comparisons of the *no change versus deteriorated* and the *improved versus deteriorated groups*. A second MLR changing the reference group from the “last” category to the “first” category (BASE=FIRST ORDER=ASCENDING) must then be run to produce the last remaining comparison, the *improved versus no change* groups. The user also has the choice to order the remaining categories of the dependent variable in “ascending” or “descending” order, via the numeric code assigned to the value label. For example, in the case illustrated here, level 2 = *deteriorated* is assigned as the reference category (the LAST category of the *cgasgroup* variable), and the remaining levels will be presented in ascending order (meaning 0 = *no change* will be followed by 1 = *improved*). It is common to see either the first or last category of the variable used as the reference category. All independent variables are then listed after the BY command. If a continuous covariate was included, it would be included after a WITH command (none is included here). Independent variables will be automatically dummy-coded by SPSS in the appropriate regression 0–1 format (SPSS Command Syntax Reference, 2005).

Lines 3 and 4 provide the criteria for estimating the MLR. All items are the default settings. Note that DELTA can be changed from 0 to .05 to stabilize the algorithm because of 0 frequencies in a small number of cells (Hosmer & Lemeshow, 2000; Tabatchnick & Fidell, 2007). This may be found to be necessary if the initial MLR results in an SPSS warning message stating there are 20% or more of the subpopulations (cells) with 0 frequencies. Typically, 20% has been set as the maximum number of cells with a small cell size (Tabatchnick & Fidell, 2007). This commonly occurs because of the geometrically increasing number of cells with each level of every independent variable laid out in a contingency table.

Line 5 signifies that a main effects MODEL will be estimated. This means no interactions will be analyzed unless they are manually entered. If the MODEL command is present, the STEPWISE command (not shown) is ignored.

The backward and forward stepwise variable selection can be utilized in MLR. Line 6 includes the INTERCEPT in the model. This is the default, and refers to the prediction of the model without any independent variables.

Line 7 produces the classification table (CLASSTABLE), the Pearson and Deviance goodness-of-fit chi-square tests (FIT), the parameter estimates for the full model using maximum likelihood estimation (PARAMETER), the pseudo R^2 (SUMMARY), the Likelihood Ratio tests for the independent variables (LRT), the case processing summary (CPS), model fitting information using the chi-square likelihood ratio test (MFI), and the AIC and BIC fit statistics (IC). Line 8 computes and saves the estimated probabilities and predicted probabilities used in the classification table.

Table 3 begins the SPSS output. The case processing summary presents the frequencies first for the dependent variable, followed by all the independent variables that were retained in the final model. The total number of valid cases used in the analysis is listed, followed by the number of cases with missing data on any one or more of the variables in the final model. The subpopulations are all the levels of the independent variables multiplied together (in this case, 2 levels \times 3 levels \times 2 levels = 12).

In order to attain stability in the statistical analysis, a standard cases-to-variables ratio is a minimum of 10 cases to every 1 variable, with a 20:1 ratio much preferred. Including all levels of each variable then, would result in a 447/10 or 45:1 cases-to-variables ratio, which should be adequate, but it will depend on how the cases are distributed across all the variables.

Model Fit

There are several ways to assess model fit in MLR. Most commonly used is the likelihood ratio test, shown in Table 4. Here, the -2 log likelihood is computed for the intercept only model, or the model without any independent variables, and the final model with all independent variables. The -2 log likelihoods for each are subtracted from one another to produce the chi-square ($133.455 - 86.675 = 46.78$). A greater amount of change between the two models

TABLE 4. Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 log likelihood	Chi-square ^a	df	Sig.
Intercept only	137.455	145.660	133.455			
Final	106.675	147.700	86.675	46.780	8	.000

^aComputed by subtracting -2 log likelihoods of final model from intercept: $133.455 - 86.675 = 46.780$. A larger chi-square suggests a greater contribution of the IV's to the DV than the intercept-only model. Statistical significance ($<.05$) suggests good model fit.

suggests a greater improvement in model fit. Significance at less than .05 suggests model fit. Here we see that the final model is significantly different from the intercept-only model ($p < .001$). Thus, the independent variables, as a group, contribute significantly to prediction of the outcome. A lower Akaike's Information Criterion for the final model compared to the intercept-only model also suggests good fit (Tabatchnick & Fidell, 2007).

Goodness of Fit

Pearson chi-square and deviance statistics assess the goodness of fit of the model and the SPSS output generates a goodness-of-fit table (output not shown here). Here, statistical significance is not desired because it would indicate a difference between the final model and a perfect model. Nonsignificance "indicates that the final model adequately duplicates the observed frequencies at the various levels of the outcome" (Tabatchnick & Fidell, 2007, p. 459).

Effect Size

SPSS generates three different pseudo R^2 summary statistics, used by some to assess model fit by determining the effect size of the model. For this analysis, pseudo R^2 statistics were as follows: Cox and Snell, .099; Nagelkerke, .120; and McFadden, .059. There is disagreement on use of the pseudo R^2 ; Tabatchnick and Fidell (2007) suggest that it approximates the same variance interpretation as R^2 in linear regression (meaning accounting for the amount of explained variance in the outcome variable), and Hosmer and Lemeshow (2000) recommend using it for model building only. It is instructive

nonetheless to become familiar with the three commonly included pseudo R^2 statistics. All of the pseudo R^2 statistics are typically much lower than the R^2 statistics in linear regression.

McFadden's is a transformation of the likelihood ratio statistic. It is computed by dividing the following:

$$\frac{1 - (\text{the ratio of the loglikelihood of the full model})}{(\text{the log likelihood of the constant only model})}$$

Values from .2 to .4 for the McFadden are considered "highly satisfactory" (Hensher & Johnson, 1981; Tabatchnick & Fidell, 2007). The Cox and Snell pseudo R^2 is also based on the log likelihoods and takes into account sample size, but it cannot achieve a maximum value of 1. Nagelkerke pseudo R^2 adjusts Cox and Snell so that a value of 1 is possible. The pseudo R^2 in the example, then, would be considered somewhat weak.

Likelihood ratio tests present the significance of the independent variables computed independently for each of the independent variables in the model (output not shown). This tests the improvement in the model fit with each of the predictor variables added (Tabatchnick & Fidell, 2007). Statistical significance ($<.05$) is desired for each of the independent variables. This was the case for the independent variables in this example.

Parameter Estimates for the Final Model

Table 5 presents the parameter estimates for the final model. The "deteriorated" group is the reference group. It is compared to the "no change" group, with parameters estimated, and the "improved" group, with separate parameter

TABLE 5. Parameter Estimates: "Deteriorated" as Reference Category

3-category change in CGAS ^a	B	Std. Error ^d	Wald	df	Sig.	Exp(B)	95% confidence interval for Exp(B) ^c	
							Lower bound	Upper bound
No change^a								
Intercept	.337	.374	.815	1	.367			
No neglect	.582	.352	2.728	1	.099	1.789	.897	3.569
Neglect	0**	.	.	0
Normal CGAS	-.225	.337	.446	1	.504	.798	.412	1.546
Clinical CGAS	0**	.	.	0
Extreme Poverty	-.678	.495	1.874	1	.171	.508	.193	1.340
Not extreme poverty	-1.193	.371	10.332	1	.001	.303	.146	.628
DK/poverty	0**	.	.	0
Improved^b								
Intercept	.855	.320	7.132	1	.008			
No neglect	1.070	.278	14.769	1	.000	2.915 ^e	1.689	5.031
Neglect	0**	.	.	0
Normal CGAS	-.862	.269	10.252	1	.001	.422 ^f	.249	.716
Clinical CGAS	0**	.	.	0
Extreme Poverty	.425	.404	1.102	1	.294	1.529	.692	3.377
Not extreme poverty	.106	.306	.119	1	.730	1.111	.610	2.025
DK/poverty	0**	.	.	0

*The reference category is: deteriorated.

**This parameter is set to zero because it is redundant.

^aThis data compares "no change" group and reference group ("deteriorated")

^bThis data compares "improved" group and reference group ("deteriorated")

^cConfidence intervals that have a smaller range and do not include "0" suggest greater precision.

^dNumerical errors are suggested if standard errors are > 2.0.

^e"Exp. B" are odds ratios for those in that level of the variable to be in the group indicated compared to the reference group. For example, those who were "not neglected" were 2.9 times as likely (more likely) to be in the "Improved" group vs. the "deteriorated" group compared to those who were not neglected.

^fThose with a normal CGAS were less likely (.42 odds ratio) to be in the "improved" group versus the deteriorated group compared to those with a clinical CGAS. If we take the reciprocal of .42, we can say that those with a normal CGAS are 2.3 times as likely to be in the deteriorated group than the improved group.

estimates. This is one of the main strengths of MLR—different estimates are computed for all paired groupings of the dependent variable. This means that different effects of particular variables within each group can be identified.

Before examining the influence of the independent variables, it is important to assess the model for possible numerical errors. One quick way is to check for standard errors that are greater than 2. This could mean high multicollinearity between independent variables. As in logistic regression, it is also a good idea to check the 95% confidence intervals. For the significant variables, smaller confidence intervals suggest greater precision of the variable. Confidence intervals containing 1 also denote a lack of significance because this would mean the range of the odds was both higher and lower than 1, making it simultaneously more and less likely to

occur (Hosmer & Lemeshow, 2000; Tabatchnick & Fidell, 2007). It is also important to account for the increased Type I error because of the large number of statistical tests being run. Rather than going with a standard $p < .05$ critical value, it is recommended that a corrected value in which .05 is divided by the total number of predictor variables be utilized (for our example, $.05 \div 3$ independent variables = .016) (Tabatchnick & Fidell, 2007). Thus, for this example, statistical significance was determined at a $p < .01$ level.

Returning to the parameter estimates for the "no change" versus "deteriorated" and "improved" versus "deteriorated" groups shown in Table 5, it was found that the "deteriorated" and "improved" groups were different on two variables, neglect and CGAS, but not poverty. The Exp(B) is the exponentiated beta, or the odds

ratio. Recall that odds ratios less than 1 indicate a lower likelihood for the event of interest; odds ratios greater than 1 indicate greater likelihood for the event of interest. Children who were not neglected were 2.9 times as (more) likely to be in the “improved” group versus the “deteriorated” group compared to children who were neglected. Children who had a normal CGAS at intake (not clinical) were .42 times as likely (about half as likely) to be in the “improved” group versus the “deteriorated” group. In other words, using the reciprocal of .42, we can state that children who had a normal CGAS at intake were 2.3 times more likely to be in the “deteriorated” group than the “improved” group (Dunn & Clark, 2001; Rosenthal, 2001). In the “no change” versus the “deteriorated” group comparison, only not being in extreme poverty significantly predicted being in the “no change” group. Those who were not in extreme poverty were less likely to be in the “no change” group versus the “deteriorated” group. Note that in a three-level independent variable, the third pairwise comparison (in this case, between the extreme poverty and not in extreme poverty groups) is not statistically tested. Therefore, in a separate analysis, this variable was recoded in the reverse, so that the extreme poverty group became the reference group. Our separate analysis with the extreme poverty group as the reference group revealed no further differences between the extreme poverty groups.

Parameter estimates for the second MLR that used the last category (no change) as the reference group were also run, but in the interests of space, are not shown here. This presents the last pairwise comparison of interest: “improved” versus “no change” (we’ve already discussed “deteriorated” versus “no change”). When we look at the model for “improved” versus “no change,” we see that the two levels of the poverty variable are significant at $p < .01$: caregiver extreme poverty and not being in extreme poverty, when each was compared to the “don’t know” cases. Children with caregivers who were in extreme poverty were 3 times as likely to be in the “improved” versus the “no change” group compared to children for whom poverty was unknown. Children not in extreme poverty were 3.6 times as likely to be in the “improved” category versus the “no change” category compared

to children for whom poverty was not known. In a separate analysis, this variable was recoded in the reverse, so that the “extreme poverty” group became the reference group. The parameter estimates revealed no difference between those known to be in poverty and those not in poverty. Neglect and having a CGAS in the clinical range at intake did not differentiate children who experienced “no change” versus children who “improved.”

Classification

The classification table is considered another indicator of the usefulness of the final model and is shown in Table 6. The observed (actual frequencies in the data) versus the predicted groupings are compared. Correctly classified cases are on the diagonal (3 cases in the “no change” group, 291 cases in the “improved” group, and 0 cases in the “deteriorated” group). Here we see that, overall, the final model accurately predicted 65.8% of the cases. However, we see that the “improved” group had a much higher level of accurate prediction at 99.7% compared to the other two groups. This is a somewhat common occurrence in MLR. For the 79 cases in the “deteriorated” group, the predicted model incorrectly classified five as “no change” and 74 as “improved.” For the 76 cases in the “no change” group, the model incorrectly classified 73 as “improved.” For the “improved” group, only one

TABLE 6. Classification

Observed	Predicted			Percent correct ^b
	No change	Improved	Deteriorated	
No change	3 ^a	73	0	3.9
Improved	1	291 ^a	0	99.7
Deteriorated	5	74	0 ^a	.0
Overall percentage	2.0	98.0	.0	65.8

^aCorrectly classified cases are on the diagonal; 3 cases in the “no change” group were correctly classified; 291 cases in the “improved” and no deteriorated cases were correctly classified.

^bOverall correct classification for all cases was 65.8%. However, correct classification was only 3.9% for the “no change” group and no deteriorated cases were correctly classified. This model then, is most useful for those whose CGAS scores improved.

case was incorrectly classified as “no change.” No cases were accurately predicted to be in the “deteriorated” group, and very few cases were accurately predicted for the “no change” group. Unfortunately, MLR generally produces the best predictions for the largest group (Hosmer & Lemeshow, 2000).

Still, the usability of this model to predict who will improve is worth pursuing. This is done by computing the proportional-by-chance accuracy rate of the classification. The chance-accuracy rate is simply the percentage categories or marginal percentages of the dependent variable shown in the case-processing summary in Table 3. These are squared and added up to get the proportional-by-chance accuracy rate of the existing data. A 25% improvement over this chance rate has been set as an acceptable standard. Using the marginal frequencies provided in the case processing summary, we get the following equation for the proportional-by-chance accuracy rate to determine a 25% prediction improvement: $(0.170^2 + 0.653^2 + .177^2 = .487)$, then $(1.25 \times .487 = .609)$. If 65.8% > 60.9%, then the model improves on chance by 25% or more and is considered adequate. Thus, 65.8% correctly classified cases exceeds the proportional-by-chance accuracy rate for this data of 60.9%, so the model has adequate accuracy. The researcher should bear in mind which groups of the dependent variable have the strongest prediction.

Strengths and Limitations of MLR

The limitations of MLR include: (a) significantly large sample sizes across all levels of the dependent and independent variable are needed for accurate estimation of parameters (Hossain, Wright, & Petersen, 2002)—this was evident in the example data set in which additional variables that would better predict the “no change” and “deteriorated” groups were needed, but because of missing data, could not be included; (b) assessing model fit is not as well developed as in linear regression methods; and (c) interpretation of the models can be difficult with more than four groups in the dependent variable (Hosmer & Lemeshow, 2000). The strength of MLR is its ability to determine differential characteristics of

client groups through estimation of coefficients for each level of the comparison of the independent/dependent variable relationships (Hosmer & Lemeshow, 2000).

Summary of MLR Process

Reviewing the entire process, like logistic regression, MLR assigns a reference group to which all other levels of the dependent variable are compared. Maximum likelihood estimation using an iterative process produces odds ratios for the independent variables for each of the paired comparisons of the dependent variable. Model fit is assessed by comparing the -2 log likelihood for the intercept-only model and the full model, and uses the chi-square statistic to determine if the improvement is statistically significant, with p values less than .05 indicating model fit. Goodness of fit can be assessed through the Pearson and Deviance chi-square tests, with p values greater than .05 signifying better fit. Three types of pseudo R^2 are produced, but these statistics are to be viewed with caution because they do not carry the same interpretation as in linear regression. Pseudo R^2 values are typically lower in MLR analyses than in linear regression. Likelihood ratio tests determine the contribution of each of the independent variables to the dependent variable, with p values lower than .05 signifying a significant contribution. Prediction accuracy of the model is assessed through the classification table. This table indicates the observed versus the predicted classification of subjects into each level of the dependent variable based on the parameter estimates. It includes what percentages of cases were accurately predicted into each level of the dependent variable (Chan, 2005; Hosmer & Lemeshow, 2000; *SPSS Command Syntax Reference*, 2005).

CONCLUSION

Multinomial logistic regression analyses have examined many areas of interest to social work practitioners. What is notable is that each of these studies was designed to generate information that could immediately be integrated into assessment, intervention, or evaluation of outcomes in clinical practice. Examples of MLR

in the social work literature include: identifying risk and protective factors for foster youth and their subsequent juvenile offending patterns (nonoffenders, early onset desisters, and chronic offenders) in which recommendations for practice differed by group (Ryan, Hernandez, & Herz, 2007); determining the likelihood of several psychiatric disorders among adults who experienced no childhood physical punishment, physical punishment only, and child abuse (Afifi, Brownridge, Cox, & Sareen, 2006); preferences for long-term care arrangements among the elderly including all formal care, all informal care, and mixed care (Min, 2005); analyzing the cognitive abilities among the elderly to determine three different ability level groups (Dodge et al., 2005); the demographic characteristics among autistic children who were not known to have experienced abuse, experienced physical abuse only, or experienced sexual abuse only (Mandell et al., 2005); the relationship between depression and suicidal ideators, suicidal attempters, and nonsuicidal adolescents (Wild et al., 2004); the prevalence of alcohol consumption in the elderly in 5-year age groups (Breslow, Faden, & Smothers, 2003); and the demographic, family, and church participation factors among African Americans (Chatters, Raylor, Lincoln, & Schroepfer, 2002).

Analyses can be conducted with sample sizes ranging from several hundred to several thousand cases, using prospective or retrospective data. The three (or more) unordered group structure can be used as a dependent or outcome variable for group classification purposes, or as a predictor variable. By devising group classifications that are of importance to clinicians, findings can be of immediate use to practice. It is the author's hope that social work researchers will be willing to learn by doing when it comes to categorical data analyses, finding creative and efficient ways to mine the data available within existing program services so that research can contribute to improved practice on behalf of clients.

REFERENCES

- Abu-Bader, S. H. (2006). *Using statistical methods in social work practice: A complete SPSS guide*. Chicago: Lyceum Books, Inc.
- Affifi, T. O., Brownridge, D. A., Cox, B. J., & Sareen, J. (2006). Physical punishment, childhood abuse and psychiatric disorders. *Child Abuse & Neglect*, *30*, 1093–1103.
- Allison, P. D. (1995). *Survival analysis using the SAS system: A practical guide*. Cary, NC: SAS Institute, Inc.
- Breslow, R. A., Faden, V. B., & Smothers B. (2003). Alcohol consumption by elderly Americans. *Journal of Studies on Alcohol*, *64*(6), 884–892.
- Chan, Y. H. (2005). Biostatistics 305: Multinomial logistic regression. *Singapore Medical Journal*, *46*(6), 259–268.
- Chatters, L. M., Raylor, R. J., Lincoln, K. D., & Schroepfer, T. (2002). Patterns of informal support from family and church members among African Americans. *Journal of Black Studies*, *33*(1), 66–85.
- Dodge, H. H., Kadowski, T., Hayakawa, T., Yamakawa, M., Sekikawa, A., & Ueshima H. (2005). Cognitive impairment as a strong predictor of incident disability in specific ADL-IADL tasks among community-dwelling elders: The Azuchi study. *The Gerontologist*, *45*(2), 222–230.
- Dunn, O. J., & Clark, V. A. (2001). *Basic statistics: A primer for the biomedical sciences* (3rd ed.). New York: John Wiley & Sons.
- Gingerich, W. J. (1990). Rethinking single-case evaluation. In L. Videka-Sherman & W. J. Reid (Eds.), *Advances in clinical social work research* (pp. 11–24). Silver Spring, MD: NASW Press.
- Goldstein, E. G. (2007). Social work education and clinical learning: Yesterday, today, and tomorrow. *Clinical Social Work Journal*, *35*, 15–23.
- Hensher, D., & Johnson, L. W. (1981). *Applied discrete choice modeling*. London: Croom Helm.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons.
- Hossain, M., Wright, S., & Petersen, L. A. (2002). Comparing performance of multinomial logistic regression and discriminant analysis for monitoring access to care for acute myocardial infarction. *Journal of Clinical Epidemiology*, *55*, 400–406.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury Thomson Learning.
- Kazi, M. A. F. (2003). *Realist evaluation in practice: Health and social work*. London: Sage Publications.
- Kirk, S. A. (1990). Research utilization: The substructure of belief. In L. Videka-Sherman & W. J. Reid (Eds.), *Advances in clinical social work research* (pp. 233–250). Silver Spring, MD: NASW Press.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.
- Mandell, D. S., Walrath, C. M., Manteuffel, B., Sgro, G., & Pinto-Martin, J. A. (2005). The prevalence

- and correlates of abuse among children with autism served in comprehensive community-based mental health settings. *Child Abuse & Neglect*, 29, 1359–1372.
- Min, J. W. (2005). Preference for long-term care arrangement and its correlates for older Korean Americans. *Journal of Aging and Health*, 17(3), 363–395.
- Reid, W. J., & Fortune, A. (1992). Research utilization in direct social work practice. In A. Grasso & I. Epstein (Eds.), *Research utilization in the social services* (pp. 233–250). New York: Haworth.
- Rosenthal, J. A. (2001). *Statistics and data interpretation for the helping professions*. Belmont, CA: Brooks/Cole Thomson Learning.
- Ryan, J. P., Hernandez, P. M., & Herz, D. (2007). Developmental trajectories of offending for male adolescents leaving foster care. *Social Work Research*, 31(1), 83–93.
- Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., et al. (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry*, 40, 1228–1231.
- Simpson, G. A., Williams, J. C., & Segall, A. B. (2007). Social work education and clinical learning. *Clinical Social Work Journal*, 35, 3–14.
- SPSS Command Syntax Reference*. (2005). Chicago: SPSS, Inc.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed., pp. 481–498). Boston: Pearson Education, Inc.
- Thomas, E. J., & Rothman, J. (1990). An integrative perspective on intervention research. In J. Rothman & E. J. Thomas (Eds.), *Intervention research: Design and development for human service* (pp. 3–23). New York: Haworth Press.
- Thyer, B. A. (2007). Social work education and clinical learning: Towards evidence-based practice? *Clinical Social Work Journal*, 35, 25–32.
- Wild, L. G., Flisher, A. J., & Lombard, C. (2004). Suicidal ideation and attempts in adolescents: Associations with depression and six domains of self-esteem. *Journal of Adolescence*, 27(6), 611–624.
- Wood, K. M. (1990). Epistemological issues in the development of social work practice knowledge. In L. Videka-Sherman & W. J. Reid (Eds.), *Advances in clinical social work research* (pp. 373–390). Silver Spring, MD: NASW Press.