Educational and Psychological Measurement
Volume 66 Number 2
April 2006 228-239

10.1177/0013164405278580
http://epm.sagepub.com
hosted at
http://online.sagepub.com

Loss of Power in Logistic, Ordinal Logistic, and Probit Regression When an Outcome Variable Is Coarsely Categorized

Aaron B. Taylor
Stephen G. West
Leona S. Aiken
Arizona State University

Variables that have been coarsely categorized into a small number of ordered categories are often modeled as outcome variables in psychological research. The authors employ a Monte Carlo study to investigate the effects of this coarse categorization of dependent variables on power to detect true effects using three classes of regression models: ordinary least squares (OLS) regression, ordinal logistic regression, and ordinal probit regression. Both the loss of power and the increase in required sample size to regain the lost power are estimated. The loss of power and required sample size increase were substantial under conditions in which the coarsely categorized variable is highly skewed, has few categories (e.g., 2, 3), or both. Ordinal logistic and ordinal probit regression protect marginally better against power loss than does OLS regression.

Keywords: *statistical power; variable categorization; OLS regression; logistic regression; probit regression*

This article investigates the loss of power in a regression analysis that occurs when a continuous dependent variable is coarsely categorized into a small number of categories. Coarse categorization typically occurs in one of two ways. In the first, a variable is initially measured using a continuous scale or one having a large number of ordered categories and is then coarsely categorized for purposes of analysis. This kind of coarse categorization frequently occurs in psychiatric settings, where an effectively continuous scale is used to measure a clinical variable, and it is then dichotomized into "case" or "noncase" (Streiner, 2002). In the second, a theoretically continuous variable is measured using only a small number of ordered categories. This practice is common in social and in developmental psychology, in which attitudes and other personal characteristics may be measured using a five-category or even a three-category scale.

Authors' Note: Please address correspondence to Aaron B. Taylor, Arizona State University, Dept. of Psychology, Tempe AZ 85287-1104, USA; tel. no.: 480-965-0914; fax no.: 480-965-8544; e-mail: Aaron.Taylor@asu.edu.

Coarsely categorized dependent variables are frequently modeled using either ordinary least squares (OLS) regression or extensions of binary outcome regression models such as ordinal logistic and ordinal probit regression (e.g., Long, 1997). Therefore, we consider the question of which of these models minimizes the loss of power when modeling a coarsely categorized outcome variable versus the normally distributed continuous representation of that same outcome variable.

Much past research has demonstrated the deleterious effects of coarse categorization. Cohen (1983) studied the effect of dichotomizing one or both of a pair of normally distributed variables on their correlation. He found that the magnitude of the correlation was attenuated, with the amount of attenuation increasing as the dichotomization moved away from an even split of the cases into two categories. Bollen and Barb (1981) studied correlated pairs of normally distributed variables, which they coarsely categorized into varying numbers of evenly distributed categories. They found that categorization into fewer categories caused a greater decline in the magnitude of the correlation. MacCallum, Zhang, Preacher, and Rucker (2002) used a series of examples to demonstrate the many negative effects of dichotomization, such as decreased effect size in OLS regression and the introduction of spurious effects in OLS regression analysis with more than one predictor variable. Krieg (1999) presented formulas for calculating biases in means, variances, and covariances that arise when continuous variables are coarsely categorized.

This article considers a very practical question that has been largely left unaddressed by previous research: How much power is lost when the *outcome variable* (dependent variable) is coarsely categorized in a regression context? Most prior work has focused on the effects of coarse categorization on measures of effect size (e.g., correlation) or on the effects of coarse categorization of predictor rather than outcome variables (MacCallum et al., 2002). Presentations of alternative regression models for ordinal dependent variables (Cohen, Cohen, West, & Aiken, 2003; Hosmer & Lemeshow, 2000; Long, 1997) have focused on the statistical properties of these models, for example, ordinal logistic regression when errors follow the logistic distribution (Hosmer & Lemeshow, p. 299; Long, 1997, pp. 119-120) rather than the normal distribution assumed in OLS regression. These presentations have also focused on the varieties of models that can be employed for ordinal outcomes, such as continuation categories, adjacent categories, and proportional odds models (Long, 1997). However, these presentations have not alerted researchers to the effects of using coarsely categorized outcome variables on statistical power.

Method

Overview

The present Monte Carlo study has two purposes. First, we investigate the size of the power loss for different types of coarse categorization of the outcome variable for each of three regression models: OLS, ordinal logistic, and ordinal probit. For each model, under the same types of coarse categorization, we also estimate the Type I error rate to check that it was controlled at the nominal level. Second, we estimate the increase in sample size for each model that would be required to achieve the same power as would have been achieved if a normally distributed continuous outcome variable had been used in an OLS regression model. In overview, the study consists of five major steps.

Step 1: Simulation of populations with continuous predictor and outcome variables. Populations of 1 million cases were simulated in which there existed a known linear relation between each of two predictors, labeled X_1 and X_2, and a continuous outcome variable, labeled Y (see the Population Simulation section below for details). In the power estimation conditions, the value of the regression coefficient of each predictor was chosen so that the probability of rejecting the null hypothesis of no effect of each predictor (i.e., the power) would be .80, given a sample size of 200 in an OLS regression model. In the Type I error estimation conditions, the regression coefficients were set equal to zero to make the null hypothesis true. The sample size was chosen to approximate the median sample size (175) found by Jaccard and Wan (1995) in their survey of multiple regression analyses reported in American Psychological Association journals in 1992.

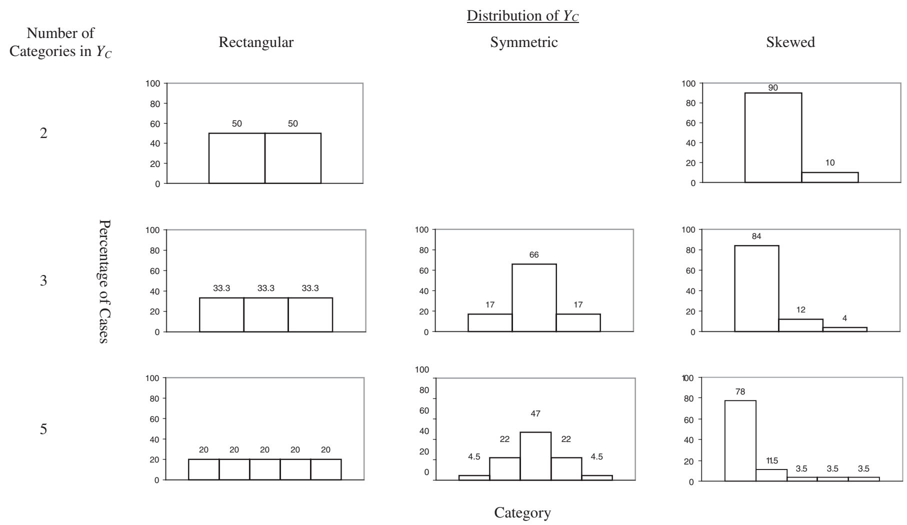
Step 2: Coarse categorization. The continuous outcome variable was coarsely categorized in the population, with the new variable being labeled Y_C. The distribution of Y_C following coarse categorization differed depending on the condition, as described below.

Step 3: Estimation of statistical power. We estimated the power to detect the effect of each predictor on Y_C by drawing 1,000 samples of size 200 from the population. Y_C was predicted as a function of X_1 and X_2 in each sample using each of the three regression models. For each model, we calculated the proportion of the 1,000 samples for which the null hypothesis of no effect in the population ($\beta = 0$) was rejected, separately for each predictor.

Step 4: Estimation of Type I error rate. We estimated the Type I error rate in each null hypothesis true condition in the same way as the power was estimated. One thousand samples of size 200 were drawn from the simulated population, the three regression models were estimated for each sample, and rejection rates of the null hypothesis were tabulated for each predictor.

Step 5: Estimation of required sample size. For the power estimation conditions, we estimated the sample size required to achieve .80 power to detect the effect of each predictor on Y_C, separately for each model. This was done once for each predictor by repeatedly drawing sets of 2,000 samples of varying size from the population until a sample size was found for which the null hypothesis of no effect was rejected for the predictor in 80% of the samples.

Figure 1
Distributions of Coarsely Categorized Outcome Variable Y_C Examined in the Study



Note: The symmetric condition is an attempt to match the skewness and kurtosis of a normal distribution, so this condition could not be implemented for the two-category condition. For the rectangular conditions, skewness = 0 and excess kurtosis = –2.00, –1.50, and –1.31 for two, three, and five categories, respectively. For symmetric conditions, skewness = 0 and excess kurtosis = –0.06. For skewed conditions, skewness = 2.67, 2.45, and 2.49, and excess kurtosis = 5.11, 5.17, and 5.39 for two, three, and five categories, respectively.

Design

The two manipulated factors of central interest in the study were the number of categories into which Y_C was categorized and the shape of the distribution of Y_C. The number of categories was set to be two, three, or five. The shape of the distribution of Y_C was set to be uniform (rectangular), unimodal and symmetric (hereafter referred to as "symmetric"), or skewed in the population. As illustrated in Figure 1, in the rectangular conditions, the proportions of cases in each of the levels of Y_C were equal in the population. For example, in the two-category uniform condition, each category of Y_C included 50% of the cases. In the symmetric conditions, the proportions of cases in the ordered categories of Y_C were chosen so that the skewness and kurtosis of the distribution would approximate the skewness and excess kurtosis of a normal distribution in the population (skewness = 0; excess kurtosis = 0). Note that the symmetric condition and the uniform condition collapse and are identical when the number of categories is two. Finally, in the skewed conditions, the proportions of cases in the ordered categories levels of Y_C were chosen so that the skewness and kurtosis of the distribution in the population would approximate the skewness and kurtosis of a two-category variable having 90% of the cases in one category and 10% of the cases in the other (skewness = 2.67; excess kurtosis = 5.11). This shape was chosen to be representative of a relatively extreme level of skewness in educational and psychological research.

We also included a third manipulated factor in the study, the intercorrelation of the two predictors in the population, ρ_{12}, which was set to be either 0 or 0.5. These values were chosen to cover the range of intercorrelation between predictors commonly found in educational and psychological research. In the power estimation conditions, the power to detect the effect of each of the predictors on the continuous outcome variable was to be fixed at .80 in an OLS regression using the continuous outcome variable, regardless of the predictor intercorrelation. Therefore, the coefficients for the predictors and ρ^2_{multiple} (population squared multiple correlation) values used in simulating the populations were larger for conditions in which ρ_{12} was 0.5 than for conditions in which ρ_{12} was 0. This was not an issue in the Type I error estimation conditions: The coefficients and ρ^2_{multiple} were set to zero regardless of the level of ρ_{12}.

Because there was no symmetric distribution condition for the condition in which Y_C had two categories, the design was *not* completely factorial. Considering the number of categories and shape of the distribution, there were eight rather than nine conditions (see Figure 1). When crossed with the two levels of ρ_{12}, this resulted in 16 power estimation conditions and 16 null hypothesis true conditions.

The type of regression model was a within-subjects factor in the design. Within each of the conditions, three models were estimated: an OLS model, a logistic model, and a probit model.

Population Simulation

A population of 1 million cases was simulated for each condition in the study. For each case, three variables were constructed: the two predictors, X_1 and X_2, and the outcome variable, Y. Each of the predictors was constructed as the weighted sum of a variable representing variance common to the predictors, C, and a variable representing variance unique to that predictor, U_1 and U_2. The outcome variable Y was then constructed as the weighted sum of the predictors and a variable for random error, labeled E. Each of the four primary variables, C, U_1, U_2, and Y, was normally distributed with a mean of zero and a variance of one and was generated using the SAS 8.2 RANNOR function (SAS Institute, 2001). The variables of interest, X_1, X_2, and Y, were then constructed using Equations 1, 2, and 3. As these variables were sums of normally distributed variables with the same mean and variance, they were themselves also normally distributed (Stuart & Ord, 1987, Example 11.2).

$$X_1 = C\sqrt{\rho_{12}} + U_1\sqrt{1-\rho_{12}}. \tag{1}$$

$$X_2 = C\sqrt{\rho_{12}} + U_2\sqrt{1-\rho_{12}}. \tag{2}$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + (1 - \rho^2_{\text{multiple}})E. \tag{3}$$

As noted above, ρ_{12} is the specified intercorrelation between the predictors, set to 0 or 0.5, depending on the condition. In the power estimation conditions, the values of β_1, β_2, and ρ^2_{multiple} (the squared multiple correlation in the population) were selected to give power of .80 to detect the effect of each predictor on the continuous outcome variable Y when an OLS regression model was used. For conditions in which $\rho_{12} = 0$, $\beta_1 = \beta_2 = .185$ and $\rho^2_{\text{multiple}} = .075$. For conditions in which $\rho_{12} = 0.5$, $\beta_1 = \beta_2 = .200$ and $\rho^2_{\text{multiple}} = .139$. In the Type I error estimation conditions, $\beta_1 = \beta_2 = 0$ and $\rho^2_{\text{multiple}} = 0$.

Once a population had been simulated, the continuous outcome variable Y was coarsely categorized into Y_C. This was done by first placing Y scores in ascending order and then assigning values of Y_C to the cases to create the distribution of Y_C for the specified condition. For example, in the two-category rectangular conditions, the first 500,000 of the 1 million cases were assigned a value of 0, and the other 500,000 cases were assigned a value of 1.

Finding Power for *n* = 200

In each of the power estimation conditions, 1,000 random samples were drawn from the population. We employed standard simulation practice for sampling from a large defined population: Cases were drawn without replacement within a sample but with replacement across samples. Otherwise stated, a particular case could not appear more than once in the same sample but could appear in more than 1 of the 1,000 samples. For each sample, four regression models were estimated. First, the continuous outcome variable Y was regressed on the two predictors using OLS. This analysis served as a check on the simulation to verify that the power to detect the effect of each predictor on the continuous outcome variable was actually .80. The three other models regressed the coarsely categorized outcome variable Y_C on the two predictors. The models used were the OLS regression model (which became the linear probability model when Y_C had two categories), the ordinal logistic model (which became the binary logistic model when Y_C had two categories), and the ordinal probit model (which became the binary probit model when Y_C had two categories). For the OLS model, t tests were used to test the effects of the predictors. For the logistic and probit models, likelihood ratio (LR) tests were used. The LR tests were chosen over the more commonly used Wald tests because the LR tests tend to be more powerful under certain conditions (Hauck & Donner, 1977). An alpha level of .05 was used for tests in all models. We compared the different models to discover which would best protect against power loss when the outcome variable was coarsely categorized.

Finding Type I Error for *n* = 200

Type I error was estimated using the same procedure as was used to estimate power. For each Type I error condition, 1,000 samples of size 200 were drawn. The OLS, logistic, and probit models were estimated for each sample. Coefficients were tested using *t* tests for the OLS model and LR tests for the logistic and probit models. Rates of rejection of the null hypothesis for each of the coefficients by each of the models were tabulated. These empirical Type I error estimates were used to check on the models' ability to control Type I error at the nominal level when the outcome variable was coarsely categorized.

Finding *n* for Power = .80

The final step in the study was to estimate the required sample size to yield power = .80 to detect the effect of each predictor on Y_C. This was carried out in each power estimation condition, once for each of the three regression models (OLS, logistic, and probit) that used Y_C as the outcome variable and once for each of the two predictors. The required sample sizes were found by means of an iterative search procedure. On each step of the procedure, 2,000 samples of a specified size were drawn from the population for the condition. The regression model being examined (OLS, logistic, or probit) was estimated in each sample, and the result of the test of the null hypothesis that the effect of the predictor being examined (X_1 or X_2) on Y_C was recorded. If the observed power, the proportion of samples in which the null hypothesis was rejected, was within .005 of the target power of .80 (i.e., between .795 and .805), then the procedure stopped. Otherwise, the estimate of the required sample size was increased or decreased as appropriate, and the procedure continued.

In addition to empirical estimates of the required sample size to achieve power = .80, theoretical estimates were calculated for the logistic regression model for the conditions in which Y_C had only two categories. The calculations were based on Hsieh's (1989) Equation 4, using coefficient estimates for each entire population of 1 million cases for each condition. No similar calculations could be made for the conditions in which Y_C had more than two categories, as Hsieh's work applies only to the binary logistic model.

Results

Across conditions of the study, the effect of ρ_{12} was found to be negligible. This was expected given that values of the coefficients β_1 and β_2, and of the squared multiple correlation, ρ^2_{multiple}, used in the simulation were adjusted to maintain the same power regardless of the level of ρ_{12}. The results for the two simulated predictors, X_1 and X_2, were nearly identical, another expected result given that $\beta_1 = \beta_2$ in the population. Because the results did not vary across level of ρ_{12} or across predictors X_1 and X_2, all results are presented collapsing across levels of ρ_{12} for only one of the predictors, X_1.

Table 1
Power to Detect the Effect of the First Predictor, X_1, on the Continuous Outcome Variable, Y, and the Coarsely Categorized Outcome Variable, Y_C, as a Function of Number of Categories and Shape of Distribution of Y_C and the Type of Regression Model, for a Sample Size of 200

	Regression Model			
Number of Categories and Shape of Y_c Distribution	OLS Model of Continuous Y	OLS Model of Y_c	Logistic Model of Y_c	Probit Model of Y_c
Two categories				
Rectangular	.7890	.5545	.5625	.5610
Skewed	.7970	.3400	.3530	.3550
Three categories				
Rectangular	.8000	.7020	.7080	.7045
Symmetric	.7965	.6540	.6570	.6595
Skewed	.8050	.4540	.4770	.4800
Five categories				
Rectangular	.8115	.7595	.7630	.7710
Symmetric	.8050	.7490	.7505	.7600
Skewed	.7900	.4860	.5325	.5375

Note: Tabled values are proportions of samples for which the null hypothesis that the coefficient for X_1 is equal to zero was rejected ($p < .05$), for each model, averaged across levels of ρ_{12}. Populations were simulated using a linear model with the continuous outcome variable Y. The coefficient for X_1 was selected in the simulation so that the power to detect its effect on Y would be .80 if an ordinary least squares (OLS) regression model were used on samples of size 200. Therefore, values in the first column are included to serve as a check on the accuracy of the data simulation.

Type I Error Rate

Empirical Type I error rates were, across models, very close to the nominal alpha of .05. They ranged from .0425 to .0600. Only 1 of the 24 Type I error rates (resulting from 3 models by 16 conditions, reduced to 8 conditions by collapsing ρ_{12}) for X_1 fell outside critical values (.04045, .05955 for lower and upper limits, respectively; not adjusted for multiple tests) for the two-tailed test against the nominal alpha of .05. All three models therefore exhibited adequate control of the Type I error rate and could be studied for their power performance.

Estimates of Statistical Power

As mentioned above, an OLS regression model was estimated for each sample in each condition of the study, to check that the simulation had been implemented correctly. Observed power values for this model in each condition appear in the first column of Table 1. They depart only slightly from .80, indicating that the input values used in the simulation were correct.

Table 1 also gives observed power levels for detecting the effect of X_1 on Y_C using OLS, logistic, and probit models in each sample, for all conditions. The loss of power

Table 2
Required Sample Size to Achieve .80 Power to Detect the Effect of the First Predictor, X_1, on the Coarsely Categorized Outcome Variable, Y_C, as a Function of Number of Categories and Shape of Distribution of Y_C and the Type of Regression Model

Number of Categories and Shape of Y_c Distribution	Regression Model: OLS Model of Continuous Y Theoretical	OLS Model of Y_c Empirical	Logistic Model of Y_c Theoretical	Logistic Model of Y_c Empirical	Probit Model of Y_c Empirical
Two categories					
Rectangular	200	316	314	317	319
Skewed	200	605	605	608	605
Three categories					
Rectangular	200	256	—	249	248
Symmetric	200	275	—	273	270
Skewed	200	472	—	461	454
Five categories					
Rectangular	200	229	—	225	226
Symmetric	200	227	—	224	223
Skewed	200	422	—	377	370

Note: Tabled values are averaged across levels of ρ_{12}, rounded to the nearest unit. Alpha was set to .05 for all tests. Theoretical sample sizes for logistic regression are based on Hsieh (1989, Equation 4) but are available only for binary logistic regression. The population sampled from was simulated so that $n = 200$ cases are required to achieve power = .80 to detect the effect of X_1 on the corresponding continuous outcome variable Y using an ordinary least squares (OLS) regression model.

for these models of the coarsely categorized outcome variable was particularly dramatic in conditions in which the distribution of Y_C was skewed. In these conditions, power declined from .80 to .34 in the two-category conditions and to .45 and .49 in the three- and five-category conditions, respectively. Power also dropped substantially to .55 in two-category conditions in which the distribution of Y_C was rectangular. In conditions in which Y_C had three or five categories and a rectangular or symmetric distribution, the power loss was far smaller. The rectangular distributions of Y_C gave slightly greater power than the symmetric distributions, but the difference was small.

Differences across type of regression model were generally very small. As might be expected, the logistic and probit models, which explicitly account for the ordinal (noncontinuous) nature of the outcome variable, had slightly more power than did the OLS model applied to Y_C, with the difference being larger when the distribution of Y_C was skewed.

Estimates of Required Sample Size

Results for the search for the sample size required to achieve power = .80 are given in Table 2. These results parallel the tests of power given a sample size of 200 described above. The increase in required sample size was most dramatic for the skewed conditions. For the two-category skewed condition, the needed sample size more than tripled, from 200 to more than 600, when the outcome variable was coarsely categorized. For the three- and five-category skewed conditions, the sample needed to achieve power = .80 was about twice the required sample size (200) with the continuous outcome. Even in those conditions in which models of the coarsely categorized outcome variable showed the best performance, the uniform and symmetric conditions in which Y_C had five categories, the required sample size increased by 10% to 30%.

Differences in required sample size across the regression models were small or virtually nonexistent. The largest such effect was for the five-category skewed condition, in which the OLS model required approximately 13% more cases than did the ordinal logistic and probit models to achieve power = .80. The theoretical and empirical sample sizes were also extremely close to each other for the logistic regression model in the conditions where Y_C had two categories. This supports our empirical estimates of the needed sample sizes in the other conditions.

DISCUSSION

Alternative regression models that permit analysis of outcome measures with two or more ordered categories have deservedly gained popularity over the past two decades. When outcome variable distributions increasingly depart from symmetry, these models' assumptions are more often met than is the OLS regression assumption of continuous normally distributed residuals (Cohen et al., 2003; Long, 1997). These models are particularly advantageous with respect to predicted scores, which are kept within the actual range of the data. However, in the advocacy for these models, it has been easy to overlook the potentially deleterious effects on statistical power that can arise whenever a conceptually continuous variable is coarsely categorized.

This study investigated the power loss that occurs when a coarsely categorized outcome variable is modeled in place of a continuous one in regression analysis. The loss of power, and consequent increase in required sample size to regain power, were greatest when the coarsely categorized outcome variable had a skewed distribution or only two categories. The power loss for a five-category outcome variable, particularly with a rectangular or symmetric distribution, was quite small. This finding parallels the results of Muthén and Kaplan (1985), who found that maximum likelihood estimation performed adequately on ordered categorical outcome variables having at least five categories. Five ordered categories, so long as they produce a distribution that is not substantially skewed, appear to be sufficient to approach the statistical power of a continuous normally distributed outcome measure.

Much previous work on the effects of coarsely categorizing predictor variables has concluded that coarse categorization of a continuous predictor can significantly reduce power (see, e.g., MacCallum et al., 2002) and has strongly recommended against the practice. This study points toward a similar conclusion for outcome variables. Significant power loss will occur when an outcome variable is coarsely categorized when only two or three categories are used or the distribution of the cases into the categories is badly skewed. Although our simulation involved multiple regression, our conclusion applies equally to analysis of variance as it is a special case of the general linear model (Cohen, 1968).

Of course, in applied research, continuous variables may simply be not measurable. Even those variables that are considered to be continuous may in fact be measured only using ordered categories, as in the case of questionnaire rating scales (e.g., strongly agree to strongly disagree). When only a single item may be employed, designing the item to have five or more equally spaced alternatives will produce the optimal results with respect to statistical power (see Krosnick & Fabrigar, in press). It is even better to use a scale composed of multiple items that measure a single dimension, as the scale will more closely approximate a continuous underlying variable.

When the number of ordered categories is limited by either the research context (e.g., ability items that permit only correct or incorrect responses) or the population (e.g., very young children who may have difficulty responding with more than two or three ordered categories), the use of multiple items to measure the underlying construct is particularly important. Summating multiple items will produce scales for which power rapidly approaches the asymptotic level of power that we achieved in the simulation using a truly continuous normal distribution. This is true, however, only when the resulting distribution of the scale is symmetrical. If the resulting scale has a skewed distribution, the power will not quickly reach the level achieved with a continuous normal distribution. We carried out additional simulations of scales consisting of continuous normally distributed items that were dichotomized using the same skewed condition used in the main simulation (a 10% endorsement rate [skewness = 2.67]). These simulations yielded estimates of .48 power with 5 items and .54 power with 10 items, across regression models. A scale consisting of continuous normally distributed items that were categorized using the same five-category condition as the main simulation (skewness = 2.49) yielded estimates of approximately .60 power with 5 items and .68 power with 10 items, across regression models.

Even in cases in which the outcome variable is believed to represent distinct classes (e.g., depressed, nondepressed), it may be possible to develop latent class models that represent a continuous latent outcome variable (the probability that the participant is a member of the class) rather than a binary variable (the observed class of the participant; see Clogg, 1995; Muthén, 2002). In general, the more closely the outcome measure can approximate a continuous variable, the greater the statistical power of the test. When only a coarsely categorized outcome variable is available, researchers should use alternative regression models such as logistic regression with binary variables or ordinal logistic regression with a small number of ordered categories. However, researchers using these approaches should plan to achieve larger sample sizes to detect hypothesized effects. The needed sample sizes may be 2 or 3 times larger than those calculated for standard OLS regression methods for continuous variables (Cohen, 1988; Maxwell, 2000), particularly when the number of categories in the outcome variable is small or the distribution of the coarsely categorized outcome variable is skewed.

References

Bollen, K., & Barb, K. (1981). Pearson's *R* and coarsely categorized measures. *American Sociological Review*, *46*, 232-239.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the behavioral sciences* (pp. 311-359). New York: Plenum.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426-443.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*, 249-254.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Hauck, W. W., & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*, *72*, 851-853.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley.

Hsieh, F. Y. (1989). Sample size tables for logistic regression. *Statistics in Medicine*, *8*, 795-802.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, *117*, 348-357.

Krieg, E. F. (1999). Bias induced by coarse measurement scales. *Educational and Psychological Measurement*, *59*, 749-766.

Krosnick, J. A., & Fabrigar, L. R. (in press). *Designing great questionnaires: Insights from psychology.* New York: Oxford University Press.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19-40.

Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods*, *5*, 434-458.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*, 81-177.

Muthén, B., & Kaplan, D. (1985). A comparison of the methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.

SAS Institute. (2001). SAS 8.2 [Computer software]. Cary, NC: Author.

Streiner, D. L. (2002). Breaking up is hard to do: The heartbreak of dichotomizing continuous data. *Canadian Journal of Psychiatry*, *47*, 262-266.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics, Volume 1* (5th ed.). New York: Oxford University Press.