Journal of Family Psychology
2007, Vol. 21, No. 4, 726–735

0893-3200/07/$12.00 DOI: 10.1037/0893-3200.21.4.726

Rethinking How Family Researchers Model Infrequent Outcomes: A Tutorial on Count Regression and Zero-Inflated Models

David C. Atkins
Fuller Graduate School of Psychology

Robert J. Gallop
West Chester University

Marital and family researchers often study infrequent behaviors. These powerful psychological variables, such as abuse, criticism, and drug use, have important ramifications for families and society as well as for the statistical models used to study them. Most researchers continue to rely on ordinary least-squares (OLS) regression for these types of data, but estimates and inferences from OLS regression can be seriously biased for count data such as these. This article presents a tutorial on statistical methods for positively skewed event data, including Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial regression models. These statistical methods are introduced through a marital commitment example, and the data and computer code to run the example analyses in R, SAS, SPSS, and Mplus are included in the online supplemental material. Extensions and practical advice are given to assist researchers in using these tools with their data.

Keywords: count models, Poisson regression, zero-inflated models

Supplemental material: http://dx.doi.org/10.1037/0893-3200.21.4.726.supp

How are stressors related to physical and emotional abuse? What aspects of a couple's relationship are predictive of contempt and belligerence? What types of family structures and interactions promote or inhibit youth drug use? Although many of the topics that marital and family researchers study are infrequent events, it is with good reason that these events have been the focus of research: Abuse, contempt, and drug use have significant economic, psychological, and cultural costs for families and society. However, these variables have ramifications for the statistical models used to study them. Except in specially selected samples (e.g., men in treatment for domestic violence), outcomes like those noted above will have strongly skewed distributions. Many individuals will report no instances of these behaviors, suggesting that there are two types of individuals in the sample and potentially two separate processes to explain their behavior.

This article is a tutorial on statistical methods specifically designed for skewed event data, including Poisson and negative binomial regression and zero-inflated versions of each (Cameron & Trivedi, 1998; Long, 1997). Zero-inflated regressions combine two models: one that focuses on the presence or absence of the outcome and a second that models the extent of the outcome when it is nonzero. This captures what most family researchers believe about rare, highly skewed phenomena: There are two classes of individuals present that may have different covariates and psychological processes at play. With a few exceptions (e.g., Fals-Stewart, Birchler, & Kelley, 2006), these models have not been widely adopted within psychology, though they are routinely used in quantitative studies in economics, political science, and sociology. To assist readers in learning these methods, the example data along with R (R Development Core Team, 2005), SAS (Littell, Stroup, & Freund, 2002), SPSS (SPSS, 2006), and Mplus (Muthén & Muthén, 2005) codes to perform example analyses are included in the online supplemental material, though not all packages can perform all analyses. See the online documentation for details. Prior to introducing regression models for counts, we review the pitfalls of modeling skewed data using ordinary least-squares (OLS) regression.

Count Data: What's the Problem?

Commitment is the foundation of relationships, and when couples encounter problems, commitment is often questioned—frequently for self and partner. Thus, relationship researchers have studied the nature of commitment and what leads to unstable relationships and relationship dissolution (Bradbury & Karney, 2004). Imagine a study of commitment among couples seeking marital therapy in which commitment was operationalized as the steps taken toward separation and divorce based on the Marital Status Inventory (MSI; Weiss & Cereto, 1980). The MSI assesses

David C. Atkins, Travis Research Institute, Fuller Graduate School of Psychology; Robert J. Gallop, Department of Mathematics and Applied Statistics, West Chester University.

We thank Brian Baucom for helpful comments that greatly improved the article.

Correspondence concerning this article should be addressed to David C. Atkins, Fuller Graduate School of Psychology, 180 North Oakland Avenue, Pasadena, CA 91101. E-mail: datkins@fuller.edu

a spectrum of behaviors related to relationship separation (e.g., "Thoughts of divorce occur to me fairly frequently, about once a week or more"; "I have set up an independent bank account in my name as a measure of protecting my own interests."). A histogram of the MSI from a sample of couples seeking marital therapy is presented in Figure 1, along with fitted probabilities that are discussed later.[1]

In this sample of 268 individuals (2 per couple), approximately 24% did not endorse any steps taken toward separation and divorce, whereas about 76% endorsed between 1 and 11 steps (out of a possible 14). Although participants were clustered within couples, the within-cluster correlation was initially ignored. Methods to account for this correlation are discussed later in the article, but as a result, reported results are not wholly accurate due to this dependence (though see footnote 9). The histogram in Figure 1 would represent the outcome distribution in our study of marital commitment. Is OLS regression appropriate for these data? If not, what are the alternatives?

As any introductory statistics text will point out, regression analyses will provide correct inferences only when the data meet certain assumptions (i.e., independence, normality of the residuals, linearity of the relationship, homoskedasticity). Regrettably, researchers do not always assess the validity of their assumptions, and if these assumptions are not met, inferences might well be wrong. Data that violate model assumptions can produce incorrect standard errors and p values, and statistical programs do not inform the user when the reported p values are not to be trusted. The MSI data presented in Figure 1 likely violate multiple assumptions but almost certainly the assumption that the residuals from an OLS regression analysis are normally distributed.[2] What are possible remedies?

One option is to do nothing and proceed with OLS regression. Given the authors' experiences with reviewing manuscripts, this strategy is not uncommon. At times, the central limit theorem (CLT) is invoked when assumptions are not met perfectly. The CLT states that as the sample size increases, the sampling distribution of the mean (or regression coefficient) becomes normally distributed regardless of the shape of the original distribution in the sample (e.g., see Cohen, Cohen, West, & Aiken, 2003). However, there are problems with invoking the CLT with data such as those presented in Figure 1: (a) it is rarely clear how big a sample size is big enough to assure that the CLT protects against Type I errors, and (b) Wilcox (2005) and others have convincingly shown that power to detect true effects plummets as assumptions are violated.

Another common remedy is to seek a transformation of the outcome variable that leads to normally distributed residuals. Common transformations include the square root, natural log, and inverse transformations (Cohen et al., 2003). Yet transformations are not without problems for infrequent event data. First, it is common to find a large proportion of the data stacked at zero, like that in Figure 1. No transformation can spread out a stack of zeros. Second, transforming the outcome fundamentally alters the structure of the residuals, and it is possible to transform the residuals to normality while concurrently violating a different assumption, such as unequal variances. Finally, King (1988) has shown that OLS regression or OLS regression using a log-transformed outcome can lead to biased and inefficient results with skewed event data. Thus, it is hazardous to use OLS regression with such data, but there are effective alternatives.

Poisson Regression for Counts

The models to be presented are specifically appropriate for data that represent counts or rates. These might include the number of contemptuous responses from a spouse within a 10-min conversation, instances of marijuana use in the last month, or incidences of physical aggression in the last year. Count and rate[3] variables share certain properties: (a) they can never be negative; (b) they are integers (i.e., whole numbers); and (c) they tend to be positively skewed. Because OLS regression uses the normal distribution as its probability model, it is fundamentally not a good fit for these types of data, as the normal distribution is symmetric and extends from negative to positive infinity.

The Poisson distribution is a much better fit for these data characteristics. Poisson regression shares many similarities with OLS regression but uses the Poisson distribution rather than the normal distribution as its probability model. Our applied introduction will not belabor the underlying mathematics and statistical theory; readers desiring further details on distributions and estimation can see Long (1997) and Cameron and Trivedi (1998). Nonetheless, there are a few critical components about the Poisson distribution (paraphrased from Long, 1997, pp. 218–219):

1. The Poisson distribution is a probability distribution for nonnegative integers.

2. The mean (or rate) of the distribution (i.e., μ) strongly controls the shape of the distribution. For

[1] The data used throughout the present article come from Christensen et al. (2004). However, due to institutional review board concerns, a small amount of random error has been added to each variable so that they are not an exact replication of the study data. Specifically, a random draw from standard normal distribution was added to each value of continuous variables. The values of the MSI were then rounded to the nearest integer. Because of this procedure, there are slight differences between the current analyses and analyses of the original data, though substantive conclusions are not affected.

[2] It is a common misconception that the outcome and continuous predictors need to be normally distributed in OLS regression, whereas the actual assumption relates to the residuals. However, the distribution of the outcome variable has a strong influence on the shape of the residuals' distribution, and it is likely that the data presented in Figure 1 would have skewed residuals.

[3] Rate within count regression is thought of as the number of counts observed within a specified time. Often, time is constant for all observations, and rate and count are synonymous, though it is possible to fit count models to rate data wherein there are different exposures (i.e., time frames). Regardless, the outcome itself is always a frequency and not a ratio of count by time.

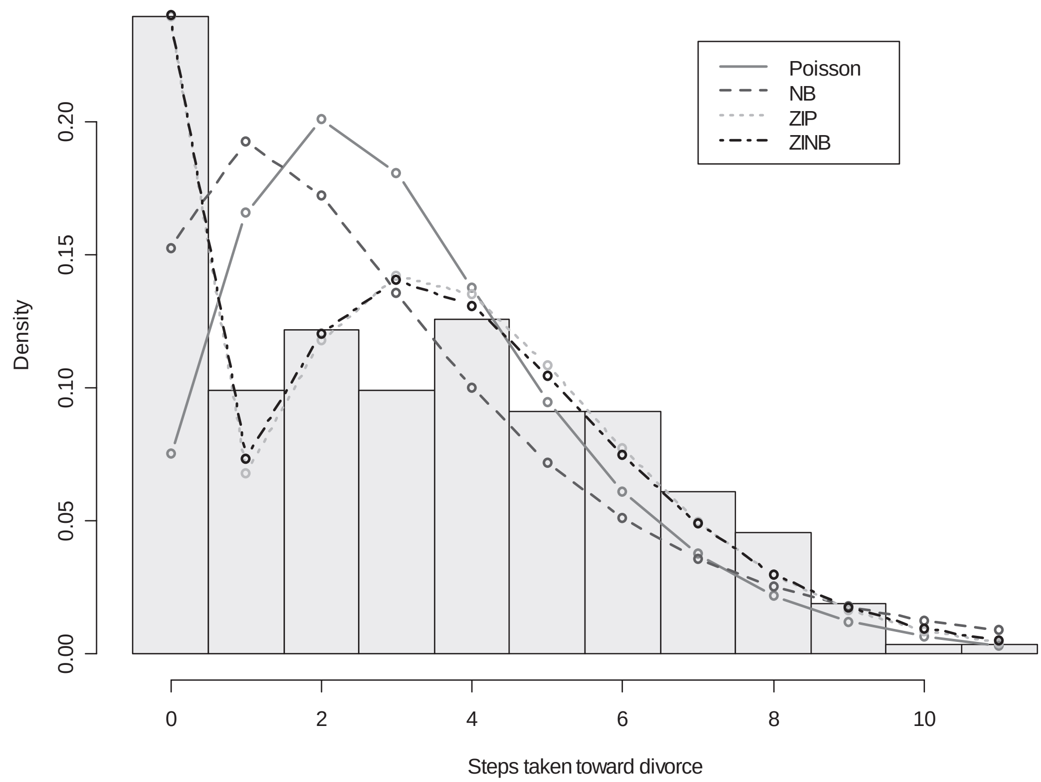

Figure 1. Histogram of Marital Status Inventory with predicted probabilities from regressions. NB = negative binomial; ZIP = zero-inflated Poisson; ZINB = zero-inflated negative binomial.

rates close to zero, the distribution is strongly, positively skewed. As the mean increases, the distribution approximates a normal distribution.

3. With the normal distribution, the mean and variance are estimated independently (e.g., two normal distributions can have the same mean but different variances). For the Poisson distribution, the variance is equal to the mean. Many times, real data do not share this property, and the variance exceeds the mean, called overdispersion, and alternative models considered below allow for overdispersion.

Each of these qualities of the Poisson distribution can be seen in Figure 2, which plots the Poisson distribution at four different means.

Even with a very small mean (e.g., $\mu = 0.5$), the Poisson distribution is never negative; rather, the distribution loads more heavily at zero. As the mean increases, the distribution becomes progressively less skewed and begins to appear normally distributed by the time the mean reaches six. OLS regression becomes more appropriate as the mean of count data increases and the data are well approximated by a normal distribution.

Poisson regression is much like OLS regression, except that the conditional distribution (i.e., residuals) is assumed to follow the Poisson distribution. Another critical difference is that the predictors are connected to the outcome via a natural logarithmic transformation. In OLS regression, the motivation for transforming a dependent variable is to normalize the residuals. In Poisson regression, the log transformation guarantees that the predictions from the model will never be negative.[4] Poisson regression is part of a broader class of models called generalized linear models (GLMs; McCullagh & Nelder, 1989), including logistic regression and OLS regression as special cases. All GLMs specify a probability distribution and a link function, such as the binomial distribution and logit link for logistic regression; for Poisson regression the link function and probability distribution are the natural log and Poisson distribution.

[4] Consider that a regression model with a natural log transformed outcome,

$$\ln(\mathrm{E}[\mathrm{Y}]) = \beta_0 + \beta_1 \mathrm{X},$$

is mathematically equivalent to a regression where the predictors are exponentiated (i.e., the predictors are the powers of base e):

$$E[Y] = e^{\beta_0 + \beta_1 X}.$$

Exponentiating the regression coefficients guarantees that the predicted values of Y will be positive, as negative numbers raised to a power of e are still positive.

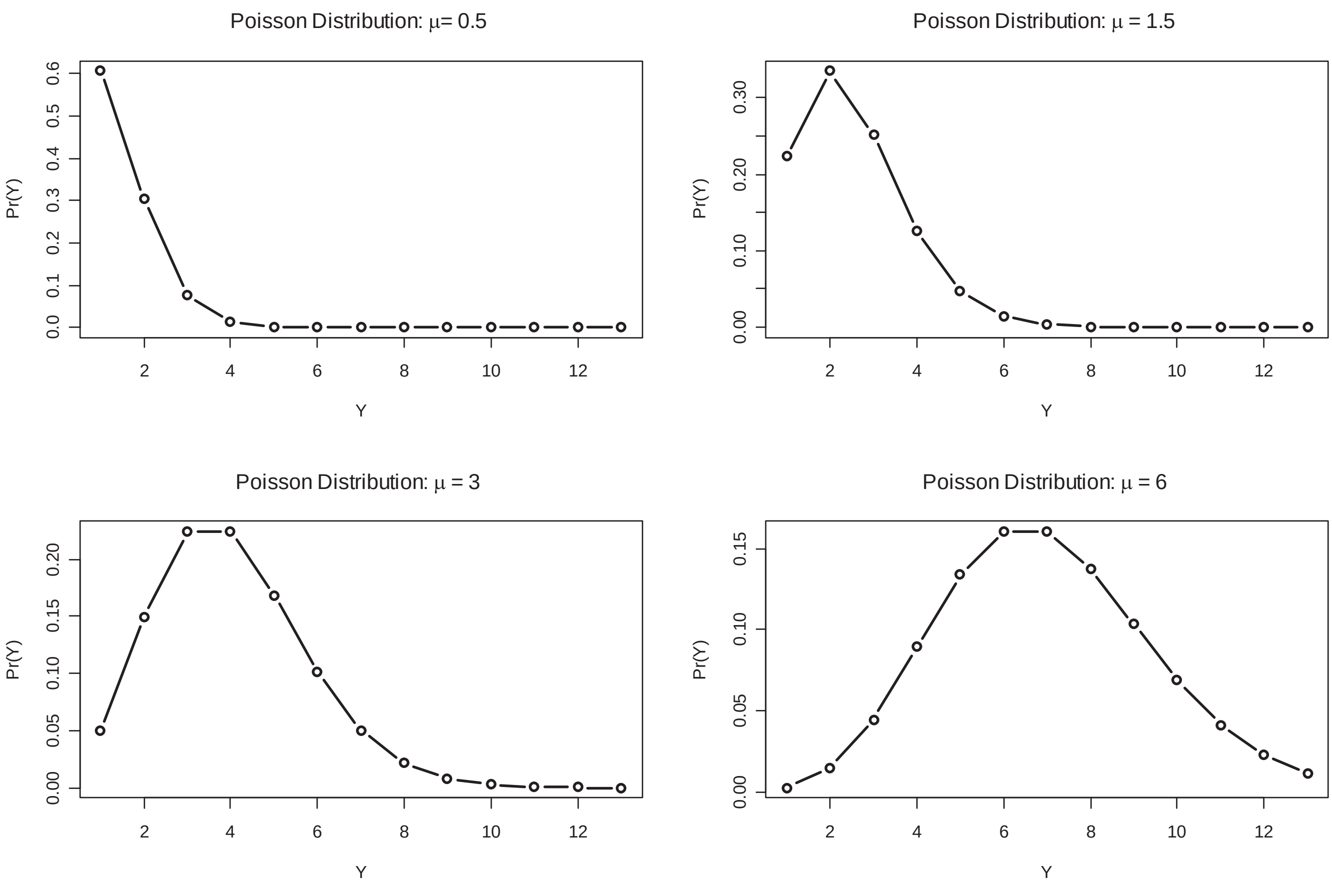

Figure 2. Shape of Poisson distribution for increasing means. Pr(Y) = probability of Y.

The overlap of count regression and GLMs is not considered further, but we did want to draw the connection for readers who might be familiar with GLMs.

Finally, estimation with Poisson regression is somewhat different from classical statistics (e.g., regression, *t* test, analysis of variance) that rely on the OLS criterion: Coefficients minimize the squared error of the regression equation. Poisson regression and the other models presented in this article rely on an alternative estimation strategy called maximum likelihood (ML).[5] ML seeks to answer the following question: What values of the regression coefficients are most likely to have given rise to the data?

ML estimation focuses on a likelihood function that describes the probability of observing the data as a function of a set of parameters. For Poisson regression, the Poisson distribution is used as the probability model, and the regression coefficients define the parameters that specify the mean structure of the data (i.e., μ in the Poisson distribution). The goal is to find the estimates of the regression coefficients that maximize the likelihood function (i.e., those values that are most likely to have given rise to the observed data). This can be accomplished by setting the first derivative of the log-likelihood[6] equation equal to zero and solving for the regression coefficients, and the final value of the log-likelihood is routinely reported in statistical output.

In most practical cases, finding ML parameters requires iterative algorithms, which adds an extra layer of complexity to these models. In particular, complex models involving many parameters and small sample sizes may prevent an algorithm from converging (i.e., finding unique parameter estimates). Finally, the results of the ML estimation yield asymptotic standard errors for the regression coefficients, and the final log-likelihood value can be used to compare the relative fit of nested models, which is described more fully later in the text.

In modeling marital commitment by means of Poisson regression, the following predictors are examined: marital satisfaction as measured by the Dyadic Adjustment Scale (DAS; Spanier, 1976), problems with affective communication (AFC) and sexual dissatisfaction (SEX) as measured by the Marital Satisfaction Inventory—Revised (Snyder, 1997), gender, and whether or not one of the partners

[5] Although ML is a fundamentally different approach to estimation than OLS, both approaches yield the same answers in the case of regression with normally distributed errors.

[6] Taking the log of the likelihood equation changes it from a multiplicative to an additive function, rendering the subsequent mathematics simpler.

Table 1
Summary of Poisson and Negative Binomial Regressions for Variables Predicting Marital Status Inventory

Variable	Poisson			Negative binomial		
	B	*SE B*	*Z*	*B*	*SE B*	*Z*
Intercept	1.13	0.05	21.96**	1.12	0.08	14.57**
DAS	−0.02	0.00	−6.14**	−0.02	0.00	−4.29**
AFC	0.02	0.01	3.41**	0.02	0.01	2.16*
SEX	0.01	0.01	1.53	0.01	0.01	1.02
Gender	−0.21	0.07	−2.97**	−0.20	0.11	−1.83
Infidelity	0.53	0.09	5.81**	0.52	0.15	3.44**
DAS × Infidelity	0.02	0.01	3.78**	0.02	0.01	2.35*

Note. $N = 263$. DAS = Dyadic Adjustment Scale; AFC = affective communication; SEX = sexual dissatisfaction. Couples without infidelity and women are coded as 0.
* $p < .05$. ** $p < .01$.

reported an extramarital affair. Finally, exploratory analyses revealed an interaction between marital satisfaction and infidelity, which is included in the model. Continuous measures were centered around their means prior to entry as predictors, binary predictors were dummy-coded with wives, and no infidelity categories were coded as zero. Results of the Poisson regression can be found in Table 1. In the present article, this model of commitment serves to introduce the statistical models, though it is prima facie a reasonable group of predictors.

As in OLS regression, each coefficient reports the effect on the outcome for a one-unit change in that predictor; however, there is a natural log link function, and coefficients are on the log scale. Effects need to be exponentiated (i.e., calculated with the inverse link function) to get estimates on the original scale of the outcome. For example, because of the coding and centering of predictors, the intercept in the model represents predicted steps toward separation and divorce for a woman who did not have an affair and who had average levels of marital satisfaction, sexual dissatisfaction, and affective communication in the sample. By transforming the intercept coefficient via $e^{1.13}$ we find that the model estimates 3.1 steps toward separation and divorce for this combination of predictors.

For the two dummy variables (i.e., gender and infidelity), the coefficients represent the change in predicted steps toward divorce from the category coded zero to the category coded one. By exponentiating each (i.e., $e^{-0.21} = 0.8$ and $e^{0.53} = 1.7$), we see that there is a small reduction for men and a modest but significant increase for those who have had affairs, though the infidelity predictor must be interpreted in light of its significant interaction with marital satisfaction. However, this method of interpretation does not directly extend to continuous predictors. The coefficients represent linear effects on the natural logarithm scale of the outcome. If we back-transform the coefficients and interpret them as in OLS regression (i.e., as the expected increase in the outcome for a one-unit change in the predictor), the coefficients would be interpreted as if they represented a linear effect on the original scale of the outcome. The crux of the matter is that because of the log link, Poisson regression is inherently nonlinear on the original scale of the outcome. Thus, continuous predictors in Poisson regression require a bit more work to interpret accurately, but three strategies aid interpretation (Long, 1997).

A regression equation is a type of prediction equation, and to interpret Poisson regression models, predictions can be generated over specific ranges of the predictors with the help of the regression equation. This strategy is also useful in OLS regression, particularly with nonlinear effects such as polynomials or interaction effects (Cohen et al., 2003), and here it focuses on the effects of marital satisfaction and infidelity and their interaction. Equation 1 presents the estimated regression equation:

$$\begin{aligned}\ln(\mathrm{E}\,[\mathrm{MSI}]) = 1.12 + 0.02(\mathrm{DAS}) + 0.02(\mathrm{AFC}) \\ + 0.01(\mathrm{SEX}) - 0.21(\mathrm{Gender}) + 0.53(\mathrm{Infidelity}) \\ - 0.02(\mathrm{DAS} \times \mathrm{Infidelity}) \quad (1)\end{aligned}$$

To generate predictions each variable in the equation must be replaced by a numeric value. The nonfocal continuous predictors (i.e., SEX and AFC) can be set to their mean values (i.e., zero because they were centered), and categorical predictors are set to their reference value (i.e., gender is set to zero, which represents wives). For the focal predictors, values across the middle 95% of our continuous predictor are generated to avoid extreme outliers (i.e., consecutive DAS values from 52 to 108 on the original scale). The DAS values are repeated: One set corresponds with the infidelity predictor equal to zero (i.e., no infidelity), and the second set of values corresponds to the infidelity predictor equal to one (i.e., infidelity). Finally, the interaction term is created by multiplying the column of DAS values with the column of infidelity values. These assorted variable values can be combined in a new, prediction data set that has six columns corresponding to our six predictors and 114 rows.[7] The three columns representing our control variables (i.e., gender, SEX, AFC) are filled with zeros. The DAS has values 52 through 108 that are repeated, which is the reason for 114 total rows. The first 57 rows have the infidelity

[7] Depending on the software, it may be necessary to generate a columns of 1s to represent the intercept as well. Moreover, some software (e.g., R) has built-in functions to generate combinations of predictors such as those described in the text, making this process significantly easier. See the online complements for an example.

predictor set to zero, and the next 57 rows have it set to a value of one.

Multiplying the variable values in each row by their corresponding regression coefficients generates predicted values for each combination of predictors that we specified in our new data set. Because of the log link, the column of predicted values is exponentiated to transform values to the original scale of the outcome, and the slopes for marital satisfaction by infidelity can be plotted or summarized in a table. Figure 3 shows the resulting plot along with 95% confidence bands, which immediately reveals the nature of the interaction.

For couples without affairs, marital satisfaction is negatively related to steps toward divorce; however, when an affair has occurred, there is essentially no relationship between marital satisfaction and steps toward divorce. The trauma of infidelity appears to push these couples closer to divorce regardless of the quality of the overall relationship. The nonlinearity of the model can also be seen in Figure 3; notice that the marital satisfaction slope for the nonaffair couples is not entirely linear. The log transformation has the strongest, nonlinear effects close to predicted values of zero, and regression lines approach zero asymptotically. Thus, the gentle slope of marital satisfaction for noninfidelity couples barely shows nonlinearity, but predictors with steeper slopes and predicted values close to zero can show strongly nonlinear effects.

A second interpretive strategy is to use the regression equation to provide predicted values for discrete combinations of predictors. Instead of generating continuous predictions along the range of predictors, specific values for individual predictors are specified (e.g., 5th and 95th percentiles of each continuous predictor, whereas other predictors are set at reference values). This strategy is virtually identical to our previous approach, except that discrete as opposed to continuous predictions are generated. For continuous variables this strategy is most appropriate when there are clear cutoffs or benchmarks (e.g., 97 on the DAS is generally considered to separate distressed from nondistressed couples).

The previous two strategies could be used with any regression model (e.g., OLS regression, logistic regression, multilevel models), whereas the third strategy is specific to Poisson regression. Because the predictors in Poisson regression are exponentiated, the Poisson regression model can be shown to be a multiplicative model as opposed to an additive model (see Long, 1997, pp. 224–226). One consequence of this is that using a simple transformation allows us to interpret regression coefficients in the Poisson model as the percentage change in the expected counts:

$$100(e^{\beta\times\delta} - 1)$$

where β is the regression coefficient from the Poisson regression and δ is the units of change in the predictor (e.g., for one unit of change in the predictor, $\delta = 1$). For example, the coefficient for AFC is 0.02 and is significantly related to the MSI. For each unit of change in AFC, the regression model predicts

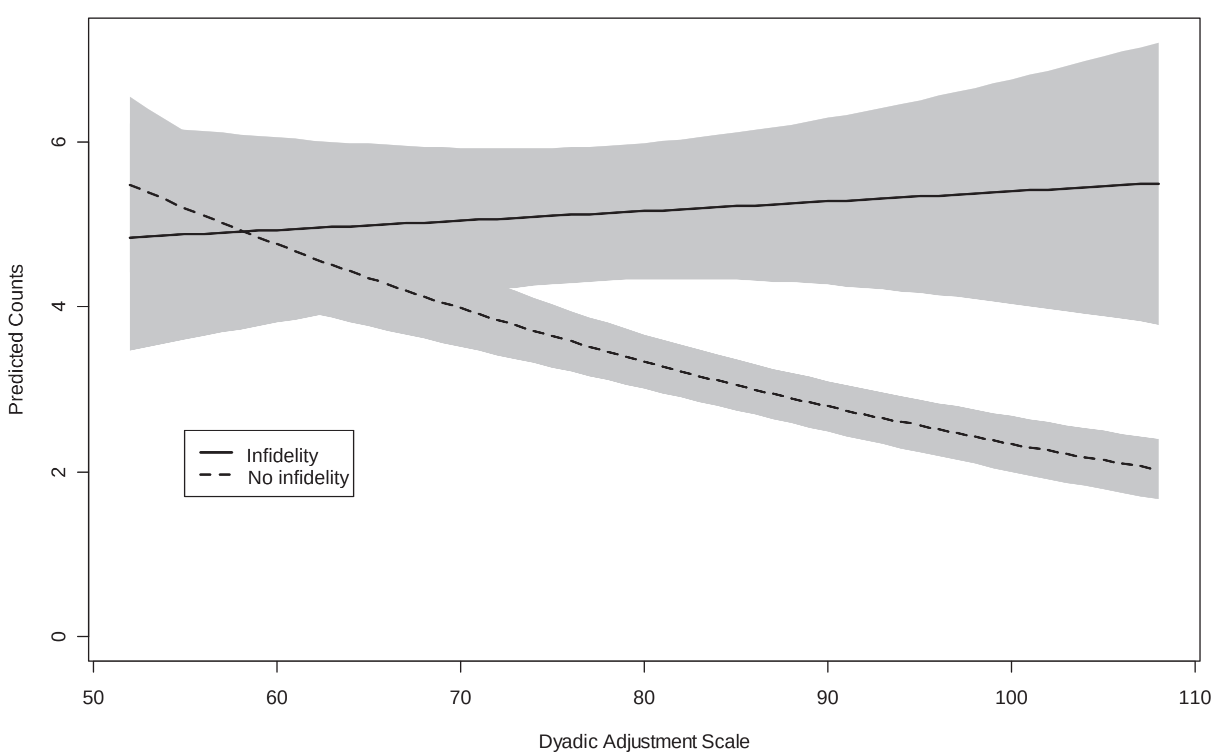

Figure 3. Predicted regression lines and confidence regions for the Infidelity × Dyadic Adjustment Scale interaction.

$$100(e^{0.02\times 1} - 1) = 2.09,$$

which is just over a 2% increase in the expected steps toward divorce. Alternatively, a standard deviation of change in AFC ($SD = 6.85$) leads to a 15.2% expected increase in the outcome. The percentage change formula can be a quick way to interpret the effects of a Poisson regression.

An important phase of any regression analysis is to assess how well the model fits the data, including residual diagnostics, influential data points, and possible nonlinearity in the predictors. Many of the common tools to assess OLS regression have been extended to Poisson regression (see Cohen et al., 2003; Fox, 1997), including standardized residuals, influence diagnostics (e.g., leverage, Cook's distance), and predictor nonlinearity (e.g., added-variable and component plus residual plots). Because of space constraints and because these tools are used in a similar fashion to their use in OLS regression, we do not present their application in the text. The online material has examples of their use.

A final tool for assessing a Poisson regression is to examine the percentage of predicted counts at each value of the outcome. How well does the model recreate the outcome distribution? Figure 1 includes predicted counts from the Poisson regression. The Poisson regression reproduces the general shape of the upper tail of the outcome distribution but sharply overpredicts the number of counts in the 1–3 range of the outcome and drastically underpredicts the zeros in the distribution. In the present data, the mean is 3.2, whereas the variance is 7.2, and the data are overdispersed relative to the Poisson distribution. With Poisson regression, the only way to increase the heterogeneity of the outcome is by adding predictors, which will have limited impact in fitting a distribution such as that in Figure 1. The limitations of the Poisson model for overdispersed data lead directly to the negative binomial regression model.

Negative Binomial Regression

The negative binomial (NB) regression model is a direct extension to the Poisson model that allows for overdispersion. In the Poisson regression model, the dispersion parameter connecting the mean and variance is fixed at one; the NB regression model is simply a Poisson regression that estimates the dispersion parameter, allowing for independent specification of the mean and variance. Because the only difference between the Poisson and the NB lies in their variances, regression coefficients tend to be similar across the two models, but standard errors can be very different. When the outcome variable is overdispersed relative to the Poisson distribution, standard errors from the NB model will be larger but more appropriate. Thus, *p* values in Poisson regression are artificially low, and confidence intervals are too narrow in the presence of overdispersion.

The earlier regression of the MSI was refit by means of NB regression, and results are found adjacent to the Poisson results in Table 1. However, is the NB model a better representation of the data relative to the Poisson? In fitting the NB model, the dispersion parameter is estimated along with its standard error, which is asymptotically normally distributed. Thus, the significance of the dispersion parameter divided by its standard error can be tested by means of the standard normal distribution; a one-tailed test is appropriate, as the dispersion can never be negative.[8] In the present data, there is strong evidence for overdispersion and the superiority of the NB model ($\theta = 2.66$, $SE = 0.55$, $Z = 4.85$, $p < .01$). Because the Poisson and NB models are nested (i.e., the Poisson model equals an NB model with dispersion fixed at one), a deviance test can also be used to test whether the NB model is a superior fit. Deviance tests are common in statistical models that are estimated via ML and compare two nested models. The test statistic is twice the difference in log-likelihoods between the two models, which is distributed as a chi-square random variable with degrees of freedom equal to the difference in number of parameters between the two models (e.g., if there were two additional predictors in Model 2 vs. Model 1, the deviance test would use a chi-square distribution with two degrees of freedom). Similar to the direct test of dispersion, the deviance test also strongly supports the NB model, $\chi^2(1, N = 256) = 61.4$, $p < .01$.

The coefficients in Table 1 are almost identical to those we saw earlier from the Poisson regression model, but the standard errors from the NB regression model are all approximately 1.5 times larger than the standard errors in the Poisson model, which follows from the variance of the NB distribution and significant dispersion. The main effect of the DAS (which represents the slope of noninfidelity couples because of the interaction and coding of infidelity), AFC, infidelity, and the interaction between DAS and infidelity remain significant at the conventional $p < .05$ criteria.

Figure 1 presents the fitted counts from the NB model. The greater variance of the NB model allows the curve of the predicted counts to be flatter and closer to the actual distribution, demonstrating the superior fit of the NB model relative to the Poisson. At the same time, the NB model predicts approximately 15% zeros in the distribution, still significantly below the approximately 24% in the MSI distribution. Data such as those represented in the MSI histogram appear to represent two separate processes at work: Some individuals never consider divorce at all, but of those who do consider divorce, there is a range of steps that different individuals have considered and pursued. Such two-step models of substantive phenomena motivated the development of statistical models that could capture these separate processes.

[8] The test for the dispersion parameter is biased and tends to be conservative because the null hypothesis is at the boundary of permissible values. This also applies to testing dispersion with the help of the likelihood ratio. Simulation-based methods provide more accurate tests (Faraway, 2005). Practically, significance tests that just fail to meet the .05 criterion may prove significant with a simulation-based test.

Zero-Inflated Poisson and Negative Binomial Regressions

When studying negative events there will often be a stack of zeros in the data, indicating that many people have never engaged in the behavior (e.g., abuse, drug use, spousal contempt). The zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regressions directly model the zeros in the structural (i.e., coefficients) portion of the model. ZIP and ZINB models are mixture models (Muthén & Shedden, 1999) in which the complete distribution of the outcome is approximated by mixing two component distributions. The most common approach is to assume a logistic regression model for the "zero, not zero" aspect of the outcome and either a Poisson or negative binomial distribution for the counts portion of the model (Lambert, 1992). The predictors for the two phases of the model can be different; thus, ZIP and ZINB are well suited for psychological models in which there are two processes and where the determinants of the two processes differ.

Table 2 presents the results from fitting ZIP and ZINB models to the MSI data, with the help of the same set of covariates for both the logistic and count portions of the models, though they need not be the same. There are two sets of coefficients for both ZIP and ZINB relating to the logistic and counts portions of the models. The coefficients for the logistic portion of the models are on the logit scale—that is, ln ($\pi/1 - \pi$) where π is the proportion of zeros—and predict the zeros in the distribution, whereas logistic regression typically predicts the ones in a binary outcome. In the same way that coefficients for Poisson (and NB) models need to be exponentiated to be interpreted, logistic models require an inverse link function to interpret. The most common way of interpreting logistic regression models is to exponentiate the coefficients, leading to an odds–ratio interpretation, though the inverse logit function could be used to interpret the results on the probability scale as well (Cohen et al., 2003; Fox, 1997). For example, to interpret the effect of the DAS on the odds of reporting no steps toward divorce (i.e., zero on the MSI), its coefficient is raised to the power of base *e* (i.e., $e^{0.05} = 1.05$). Thus, for each additional point of marital satisfaction as measured by the DAS the predicted odds of reporting no thoughts or steps toward divorce increase by 5%.

Odds ratios (or probabilities) can be plotted across ranges of predictors, or discrete change can be used in interpreting the logistic portion of the model, as introduced earlier for the Poisson regression model. In addition to the DAS, problems with AFC decrease the likelihood of no steps toward divorce (odds ratio = 0.94). In the counts portions of the models, gender, DAS, and infidelity all significantly affect steps taken toward divorce. Each additional DAS point decreases the expected steps toward divorce by 1.1%, or for each additional standard deviation in the DAS ($SD = 14.4$), the expected steps toward divorce decrease by 13.8%. The ZIP model finds that men report 18.2% fewer steps toward divorce as compared with women. Finally, couples in which there has been an affair report a 49.5% increase in steps toward dissolution. A contingency table of the MSI and infidelity reveals a large jump in the number of affair couples who have taken four or more steps toward divorce; not surprisingly, infidelity causes partners to strongly question their commitment and whether the relationship should continue.

Although the predictors are identical, effects are not always identical across the logistic and count portions. For

Table 2
Summary of Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) Regressions for Variables Predicting Marital Status Inventory

	ZIP			ZINB		
Variable	*B*	*SE B*	*Z*	*B*	*SE B*	*Z*
Logistic portion of model						
Intercept	−1.493	0.273	−5.463**	−1.516	0.282	−5.381**
DAS	0.050	0.017	2.896**	0.051	0.018	2.864**
AFC	−0.060	0.029	−2.113*	−0.061	0.029	−2.100*
SEX	−0.019	0.017	−1.128	−0.020	0.017	−1.140
Gender	−0.053	0.352	−0.151	−0.061	0.359	−0.170
Infidelity	−0.543	0.668	−0.812	−0.535	0.681	−0.786
DAS × Infidelity	−0.080	0.037	−2.176*	−0.081	0.037	−2.169*
Counts portion of model						
Intercept	1.387	0.055	25.399**	1.384	0.057	24.068**
DAS	−0.010	0.003	−3.432**	−0.010	0.003	−3.304**
AFC	0.010	0.006	1.551	0.009	0.006	1.465
SEX	0.003	0.004	0.802	0.003	0.004	0.731
Gender	−0.201	0.075	−2.676**	−0.203	0.079	−2.566*
Infidelity	0.402	0.093	4.342**	0.405	0.098	4.125**
DAS × Infidelity	0.008	0.005	1.553	0.008	0.006	1.468

Note. $N = 263$. DAS = Dyadic Adjustment Scale; AFC = affective communication; SEX = sexual dissatisfaction. Couples without infidelity and women are coded as 0.
* $p < .05$. ** $p < .01$.

example, AFC is significantly related to whether or not there are any steps toward divorce (i.e., logistic portion) but does not explain any of the variance in the number of steps taken toward divorce when there are some (i.e., counts portion). Infidelity and its interaction show a similar, differential association across the two portions of the model. The flexibility of ZIP and ZINB models for treating the two components—and potentially differential effects—of the outcome is one of its main attractions to applied researchers.

In comparing the ZIP and ZINB models, the coefficients are very similar, and the standard errors are a bit larger in the ZINB as compared with those in the ZIP model. Because the mixing proportion is estimated with the data and can be quite different between ZIP and ZINB models, coefficients at times can be very different across the two models. Thus, the current similarities are a by-product of the MSI data. Figure 1 includes the predicted probabilities from ZIP and ZINB models. The predicted count probabilities show the two-process nature of the ZIP and ZINB models and are much more accurate representations of the zeros in the data. Because the zeros are accounted for by the logistic portion of the model, the counts portion can more accurately reflect the nonzero distribution. Finally, the two zero-inflated models produce very similar counts, with the ZINB model showing slightly more heterogeneity than the ZIP. However, how should one decide between ZIP and ZINB models, or for that matter, between any of the four models presented thus far?

Choosing the Best Model for the Data

The ZIP model is nested within the ZINB in the same manner that the Poisson is nested within the NB (i.e., when the dispersion parameter equals 1, the ZINB reduces to the ZIP). The deviance test comparing the ZIP and ZINB fits of the MSI data shows no preference for ZINB, $\chi^2(1, N = 248) = 0.58$, $p > .5$. As can be seen in Figure 1, the ZINB appears to add little over the ZIP model. However, the Poisson and NB regressions are not nested within their zero-inflated extensions, precluding deviance tests. Vuong (1989) developed an appropriate test statistic for nonnested count models. Without going into the mathematical details, the test compares the predicted probabilities of counts for two different models, and the resulting V statistic is asymptotically normally distributed. Based on the Vuong test, the ZIP model is highly preferred to the NB regression model ($V = 3.97$, $p < .01$), but the ZINB is not preferred to the ZIP regression model ($V = 0.38$, $p = .35$). The extra dispersion parameter in the ZINB model does not noticeably increase the fit of the model.

An additional complexity in choosing the best model for the data is to ensure that the estimated parameters (e.g., coefficients) are the best estimates. The algorithms used to solve the log-likelihood equation can sometimes stop at a local maxima as opposed to the global maxima. Practically, the algorithm thinks it has found the best set of estimates, when in reality there are superior coefficient estimates. Teicher (1960) explains that mixture models involving the Poisson distribution are identifiable (i.e., unlikely to have local maxima and convergence issues), but this is not guaranteed for ZIP/ZINB models because of the stack of zeros. Two potential diagnostic methods for convergence issues are varying the initial parameter estimates or using the bootstrap. Statistical methods that require iterative fitting algorithms must make an initial guess at the parameters, which are then refined through the iterative search. If there is a problem with local maxima, different initial parameter estimates can lead to different final solutions. Thus, by providing a range of initial values, we can see whether the algorithm converges to the same solution or different solutions, suggesting a problem with local maxima (Hipp & Bauer, 2006).

A second approach is to bootstrap the ZIP/ZINB model (Grün & Leisch, 2004). The bootstrap is a statistical method that treats the sample of data as if it were the population (Efron & Tibshirani, 1993). A large number of new data sets (typically 999 or more) are created by sampling with replacement from the original data. In the present case, ZIP/ZINB models are run on each of the newly created data sets, and the cumulative results can be compared with the original data results. Most importantly for the present discussion, the average bias of the replicated results can be compared with the original results (i.e., on average, do the replicated results converge to what was found originally?). The online materials noted at the beginning of the article have an example of using the bootstrap with the ZINB fit.

Extensions and Future Directions

As noted earlier the example data come from spouses, who do not represent independent observations, though the models presented thus far assume independence of observations. However, most of these models have been extended to incorporate grouped and longitudinal data, though these statistical developments are ongoing and software options are more limited. The Poisson regression model has been extended to include random effects as a generalized linear mixed model (Molenberghs & Verbeke, 2005; Raudenbush & Bryk, 2002), which may include correlation models for repeated measures. The addition of random effects to the NB was considered by Min and Agresti (2005). Therefore, random effects can be introduced to describe clustering within couples.

Moreover, there have been recent extensions to the zero-inflated models to extend them to grouped and longitudinal data (Min & Agresti, 2005; Yau, Lee, & Carrivick, 2004; Yau, Wang, & Lee, 2003); however, the inclusion of both a mixture model and random effects leads to very complicated likelihood equations, which increase the possibility of convergence problems and false convergence (i.e., local maxima). A sample code to fit a Poisson mixed-effects model as well as mixed-effects ZIP/ZINB models are included in the online materials noted at the beginning of the article.[9] As is evident in the citations, these extensions are recent and ongoing, and new methods and algorithms will certainly be coming in the future.

[9] Moreover, the substantive findings presented in the text do not change when the spousal correlation is included in the model, though estimates are not identical.

This tutorial has introduced new methods for infrequent outcome variables that are critically important for family and marital researchers. Transformations or dichotomizing outcomes can be biased or ignore substantive characteristics of the sample. We hope that this introduction not only helps marital and family researchers learn and implement count regression models but also expands their thinking about how to map substantive hypotheses onto their data.

References

Bradbury, T. N., & Karney, B. R. (2004). Understanding and altering the longitudinal course of marriage. *Journal of Marriage and Family, 66,* 862–879.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data.* Cambridge, United Kingdom: Cambridge University Press.

Christensen, A., Atkins, D. C., Berns, S. B., Wheeler, J., Baucom, D. H., & Simpson, L. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72,* 176–191.

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

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* London: Chapman & Hall.

Fals-Stewart, W., Birchler, G. R., & Kelley, M. L. (2006). Learning sobriety together: A randomized clinical trial examining behavioral couples therapy with alcoholic female patients. *Journal of Consulting and Clinical Psychology, 74,* 579–591.

Faraway, J. J. (2005). *Extending the linear model with R: Generalized linear, mixed effects, and nonparametric regression models.* New York: Chapman & Hall/CRC Press.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Thousand Oaks, CA: Sage Publications.

Grün, B., & Leisch, F. (2004). Bootstrapping finite mixture models. In J. Antoch (Ed.), *COMPSTAT 2004—Proceedings in computational statistics: 16th symposium held in Prague, Czech Republic, 2004* (pp. 1115–1122). Heidelberg, Germany: Physica-Verlag.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11,* 36–53.

King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science, 32,* 838–863.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics, 34,* 1–14.

Littell, R. C., Stroup, W. W., & Freund, R. J. (2002). *SAS for linear models.* Cary, NC: SAS Publishing.

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

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). New York: Chapman & Hall.

Min, Y., & Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling, 5,* 1–19.

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data.* New York: Springer.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55,* 463–469.

Muthén, L. K., & Muthén, B. O. (2005). *Mplus: Statistical analysis with latent variables: User guide.* Los Angeles: Muthén & Muthén.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

R Development Core Team. (2005). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Snyder, D. K. (1997). *Marital Satisfaction Inventory—Revised (MSI-R) manual.* Los Angeles: Western Psychological Services.

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family, 38,* 15–28.

Statistical Package for the Social Sciences 15.0 for Windows [Computer software]. (2006). Chicago: SPSS.

Teicher, H. (1960). On the mixture of distributions. *Annals of Mathematical Statistics, 31,* 55–73.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57,* 307–333.

Weiss, R. L., & Cerreto, M. (1980). The Marital Status Inventory: Development of a measure of dissolution potential. *American Journal of Family Therapy, 8,* 80–86.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). New York: Elsevier Academic Press.

Yau, K. K. W., Lee, A. H., & Carrivick, P. J. W. (2004). Modeling zero-inflated count series with application to occupational health. *Computer Methods and Programs in Biomedicine, 74,* 47–52.

Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal, 45,* 437–452.

Received September 1, 2006
Revision received April 16, 2007
Accepted April 17, 2007 ■