Journal of Traumatic Stress, Vol. 21, No. 5, October 2008, pp. 448–454 (© 2008)

Poisson Regression for Modeling Count and Frequency Outcomes in Trauma Research

David R. Gagnon
Boston University School of Public Health, and Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, Boston, MA

Susan Doron-LaMarca
Behavioral Science Division, National Center for PTSD, VA Boston Healthcare System, and Boston University School of Medicine, Boston, MA

Margret Bell
Military Sexual Trauma Support Team, Department of Veterans Affairs Office of Mental Health Services, and Women's Health Sciences Division, National Center for PTSD, VA Boston Healthcare System, Boston, MA

Timothy J. O'Farrell
Families and Addiction Program, Department of Psychiatry, Harvard University Medical School, and VA Boston Healthcare System, Brockton, MA

Casey T. Taft
Behavioral Science Division, National Center for PTSD, VA Boston Healthcare System, and Boston University School of Medicine, Boston, MA

The authors describe how the Poisson regression method for analyzing count or frequency outcome variables can be applied in trauma studies. The outcome of interest in trauma research may represent a count of the number of incidents of behavior occurring in a given time interval, such as acts of physical aggression or substance abuse. Traditional linear regression approaches assume a normally distributed outcome variable with equal variances over the range of predictor variables, and may not be optimal for modeling count outcomes. An application of Poisson regression is presented using data from a study of intimate partner aggression among male patients in an alcohol treatment program and their female partners. Results of Poisson regression and linear regression models are compared.

In trauma research, as in other fields of research, the outcome of interest may represent a count of the number of incidents occurring in a given period of time. In the case of domestic violence research, for example, we may be examining the rate of aggressive acts during the interval between a baseline and follow-up interview. Other examples of posttraumatic research outcomes represented by count variables include behaviors, such as the frequency of substance abuse during an observation period, or events, such as the number of police interventions during a given time, or the number of days hospitalized in a year.

For these situations, it may be tempting to automatically apply linear regression, given that the count outcome variable appears to be continuous with a range of values supporting this observation. However, it is important to consider further, several assumptions of linear regression that are often violated with count data. Violation of these assumptions can result in (a) biased estimates of effect, which lead to incorrect measures of association; and (b) incorrect standard errors of the estimates, which lead to incorrect p-values and confidence intervals. In this article, we propose Poisson regression as an alternative approach for count outcome data in trauma studies and present an example that outlines how to apply it.

The data presented are taken from a study examining intimate partner aggression perpetration among male patients ($N = 178$) who had recently begun an alcohol treatment program and

This study was supported by funding from the Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, and by the following grants: Department of Veterans Affairs, CSR&D (Susan Doron-LaMarca, Principal Investigator), National Institute of Mental Health R03MH77907 (Susan Doron-LaMarca, Principal Investigator), and National Institute on Alcohol Abuse and Alcoholism R01AA10796 (Timothy O'Farrell, Principal Investigator).

Correspondence concerning this article should be addressed to: David R. Gagnon. Massachusetts Veterans Epidemiology and Research Information Center, VA Boston Healthcare System (151-MAV), 150 South Huntington Avenue, Boston, MA 02130. E-mail: gagnon@bu.edu.

 Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/jts.20359

their female partners. (For more details, see Mattson, O'Farrell, Monson, Taft, & Panuzio, 2008.) For this brief example, we present cross-sectional analyses of a subsample of participants ($n = 114$) at the initial baseline assessment. Consider Figure 1, which shows the frequency distribution of the reported number of physical assaults committed by male patients during the 6-month interval prior to the baseline assessment. Physically aggressive acts were measured with the 12-item physical assault subscale of the Revised Conflict Tactics Scales (CTS2; Straus, Hamby, Boney-McCoy, & Sugarman, 1996). In Figure 1, it is clear that the distribution of aggressive acts does not remotely follow a normal distribution, which is indicated by the black curve superimposed on the diagram. Instead, the distribution of data shows fundamental characteristics of a Poisson distribution. As presented in Figure 1, these data are highly skewed in the positive direction, which is expected, given the characteristics of domestic violence data. For instance, there are many participants with no aggressive acts, a modest number with infrequent acts, and a small number of participants with very frequent events. Remedying the lack of normality by logarithmic or square-root transformation of the outcome variable (Mosteller & Tukey, 1977) followed by linear regression is not recommended because the rates predicted by linear regression typically do not account for the large number of low or zero scores in the count outcome (Cohen, Cohen, West, & Aiken, 2003) and transformation most likely will not result in a normal distribution. In addition, introducing a log or square-root transformed outcome creates difficulties in interpretation of the effects of a predictor on the outcome.

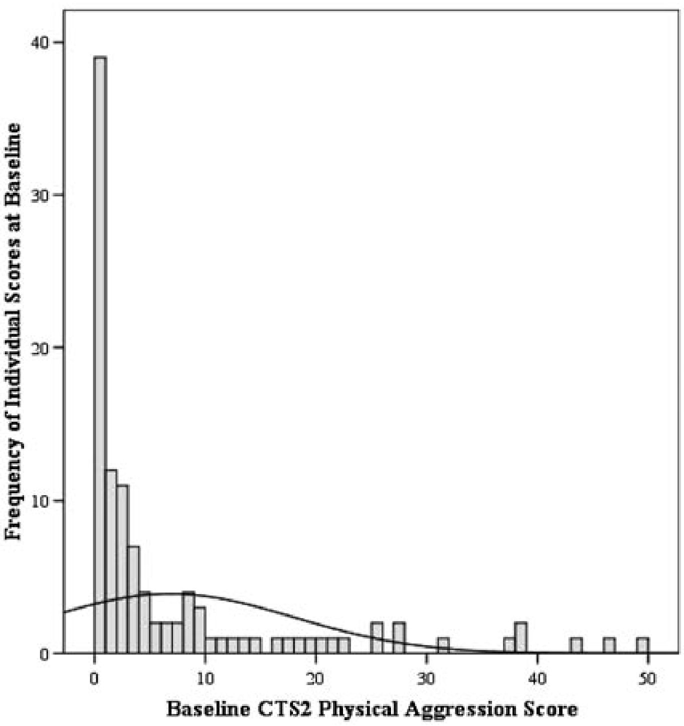

Figure 1. Frequency distribution of the combined partner report of the number of physical assaults committed by male patients during the 6-months prior to the baseline assessment as measured by the 12-item physical assault subscale of the Conflict Tactics Scales-2.

In the case of count data, attempting to predict outcomes for a given set of risk factors using linear regression can be problematic. For instance, if the number of outcome events is relatively small, a plausible risk factor profile can result in prediction of a negative number of events—an impossible situation in real life. Therefore, the regression model should be restricted to predict a positive number of events. This problem in attempting linear regression with a count outcome variable is similar to, though less severe than, the difficulty in using linear regression with a dichotomous outcome (i.e., a situation requiring logistic regression). With a dichotomous outcome, the regression model must be restricted to give predicted values between 0 and 1.00, which is a reasonable range of predicted risk. Any model that estimates the probability of an event as either negative or greater than 1.00 is incorrect.

GENERALIZED LINEAR MODELS

Generalized linear models were developed as a framework for handling such problems with data that do not meet the assumptions of simple linear regression (McCullagh & Nelder, 1989). Generalized linear models can adapt regression modeling to non-normal distributions, such as the Poisson distribution (Gardner, Mulvey, & Shaw, 1995). The Poisson distribution describes the probability of a number of events occurring during a given time interval. Interval units other than time can also be used. The Poisson distribution is defined by a rate parameter, lambda (λ), which is both the mean number of occurrences expected in a given interval, and is also equal to the variance.[1] Data that follow the Poisson distribution are not normally distributed because they are characterized by variances that are unequal over the range of predictor variables. Furthermore, the variance of the residuals about the predicted outcome depends on the value of the predicted outcome (Cohen et al., 2003).

Both Poisson regression and logistic regression are examples of generalized linear models. All regression models constituting the generalized linear model can be stated as equations that are linear in the coefficients or parameters (Cohen et al., 2003). Linear regression is also an example of a generalized linear model that does not require us to transform the outcome variable to meet model assumptions. Using a generalized linear models approach requires

[1] Example graphs of Poisson distributions for different values of lambda may be viewed with the following software applet: http://www.fortunecity.co.uk/meltingpot/ back/340/product/java/cdfdemomain.html.

us to define two factors, the link function and the distribution (Stokes, Davis, & Koch, 2001). The link function, which involves a transformation of the outcome, ensures that estimates are restricted to the appropriate range (i.e., positive for Poisson data). Defining the distribution ensures that the variances are specified appropriately (i.e., variance equal to the mean for Poisson data). In defining the link and distribution, generalized linear models do a better job of respectively ensuring that (a) the estimates are realistic and unbiased, and (b) the standard errors are appropriate for data lacking a normal distribution.

Using Link Functions to Get the Estimates Right

In a basic multiple linear regression model, the predicted value of the outcome is calculated using the following equation:

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \cdots + B_nX_n \tag{1}$$

where $X_1 - X_n$ are predictor variables of interest and $\hat{Y}$ is the predicted outcome. The estimated intercept coefficient is B_0, and $B_1 - B_n$, are the regression estimates for the strength of the effects of the predictors $X_1 - X_n$, respectively.

Without constraining any of the predictors or parameter estimates in this basic regression equation, the predicted outcome variable can take on any range of values, from negative infinity to positive infinity. Although this value range may be acceptable for linear regression, it is problematic for count outcomes, where predicted rates must be positive, and for dichotomous outcomes, where predicted risk must be between 0 and 1.00.

In the generalized linear model, the link function takes this predicted value that ranges from $-\infty$ to ∞, and transforms it to fit the range allowed by the constraints on the outcome. In the case of dichotomous outcomes for example, the link function in logistic regression works so that very large negative values from the regression equation lead to predicted risks near zero and very large values lead to predicted risks approaching 1.00. In the case of count data, we'd like to model the rate of events as a function of our regression equation. The link function in Poisson regression uses a logarithmic transformation of the rate that keeps the number of events positive, resulting in the following relationship, where N_i is the denominator used in calculating the rate:

$$\log\left(\frac{\#events_i}{N_i}\right) = B_0 + B_1X_{1i} + B_2X_{2i} + B_3X_{3i} + \cdots + B_nX_{ni} \tag{2}$$

With a little algebraic manipulation, the expression becomes

$$\log(\#events_i) = \log(N_i) + B_0 + B_1X_{1i} + B_2X_{2i} + B_3X_{3i} + \cdots + B_nX_{ni} \tag{3}$$

The $\log(N_i)$ has a special function in Poisson regression and is usually called the *offset*. It is needed because whereas a denominator is required in calculating rates, the standard errors for Poisson models are dependent on the number of events and not on the rate itself. The number of events is needed to calculate standard errors; the offset is needed to calculate the rates being compared. In a similar fashion, when creating a confidence interval for a Poisson rate, the standard error is calculated for the estimated number of events, and the upper and lower bounds of the estimated number of events is calculated. The estimate and the upper and lower bound are then divided by the offset to obtain the confidence interval for the rate. The offset can be quite flexible. For example, if one is examining the number of aggressive acts in a couple, this value of N_i would be equal to 1.00 for each couple. In person-time analysis (a favorite use of Poisson regression), the value of N_i allows for variation in event observation periods across individuals as the basis for estimating event rates. We can also predict rates of crimes in a city, where N_i is the city's population, or the rate of bird sightings in areas of a forest, where N_i is the square acres of land covered by each observer, or any other units of comparison.

Defining the Distribution to Get the Standard Errors Right

In the case of a normal distribution, the contribution of an observation to the variance is defined as $\frac{(x-\bar{x})^2}{n-1}$, which is the standard form familiar to many researchers. For a proportion, seen in a logistic model, the variance of the risk is $\frac{p(1-p)}{n-1}$. Unlike a normal distribution, which is defined by two independent parameters, i.e., a mean and a variance, a Poisson distribution is more restricted, having only one parameter, i.e., the mean. Having only one parameter makes the variance dependent on this parameter. In this case, the variance of the number of events is equal to the mean ($\bar{X}$). By specifying the appropriate distribution, we can ensure that the variances, and thus the standard errors, will be calculated correctly (Stokes et al., 2001).

Specifying Poisson Models Using SAS

Although there are many statistical packages that provide an option for Poisson regression, we used the SAS programming language (SAS Institute, 2004) to implement our example models. In SAS, PROC GENMOD is the procedure designed to handle generalized linear models. Therefore, this procedure can execute linear regression, logistic regression, Poisson regression, and several other types of analysis. By specifying the distribution and the link function, we determine the type of analysis that is performed. Below, we provide the syntax for a relatively simple example of Poisson regression in SAS.

The outcome variable, P_AGGRESSION, represents the CTS2 score, reflecting both the male and female partner combined reports of frequency of male participant physical aggression during the 6 months before the baseline assessment. Reliability of the

CTS2 has been demonstrated ranging from .79 to .95 (Straus et al., 1996). The predictor, BMIDS, represents the baseline male participant score on the Inventory of Drinking Situations (IDS; Annis, 1982). This measure indicates the extent to which a male patient drank heavily during stressful emotional situations during the 6 months prior to the baseline assessment. The IDS is a 20-item measure with responses ranging from 1 (*never*) to 4 (*almost always*) with higher scores reflecting more drinking.

For the SAS example below, we are using a data set, trauma1, which includes one observation per couple, and the variable TIME, representing the length of time the couple was under observation, as well as P_AGGRESSION and BMIDS defined above. In the case where all participants are observed for the same length of time, the TIME variable can be a constant. The syntax, data trauma2; set trauma1; creates a working or temporary data set, trauma2, from the original data set trauma1.

In creating the data set "trauma2" below, we are also defining the offset variable as the log of the TIME variable. The units for the TIME variable (e.g., days, months, years) are not important; it is only the relative size of the values that matter.

```
data trauma2; set trauma1;
ltime = log(time);
run;
```

Using the new data set, we can now run the GENMOD procedure:

```
proc genmod data = trauma2;
model p_aggression = bmids/dist = poisson link
= log offset = ltime;
run;
```

In this example, we have the standard SAS representation of the regression model (MODEL P_AGGRESSION = BMIDS) with several options after the forward slash. DIST = POISSON indicates that P_AGGRESSION is a Poisson-distributed outcome variable, and LINK = LOG specifies that the logarithmic link should be used. OFFSET = LTIME defines the offset variable to be LTIME. The offset variable is usually the logarithm of the denominator used to determine the event rates. In this case, the offset variable is the logarithm of the time under observation, resulting in rates expressed as events/unit time. The LTIME variable must be calculated in the creation of the data set as shown in the SAS syntax, e.g., LTIME = LOG(TIME).

Specifying Poisson Models Using SPSS

For researchers who may be more familiar with SPSS than SAS, below is an example of syntax for a Poisson model using SPSS version 15.0. We obtained the syntax at the following Web site: http://www.ats.ucla.edu/stat/spss/dae/poissonreg.htm (UCLA Academic Technology Services, n.d.).

```
GENLIN
p_aggression WITH bmids
/MODEL bmids
INTERCEPT = YES
DISTRIBUTION = POISSON
LINK = LOG
/CRITERIA METHOD = FISHER(1) SCALE = 1 COVB
= ROBUST.
```

The SPSS GENLIN syntax is similar to the SAS GENMOD syntax in specifying the generalized linear model distribution as Poisson and a logarithmic link function. Furthermore, the COVB = ROBUST statement in the final line of syntax corrects for underestimation of the standard errors of the estimates, which can lead to type I error as previously mentioned.

Assessing Goodness-of-Fit in a Poisson Model

Although relevant to generalized linear models in general, how well the data fits a Poisson model is of great importance. Many researchers are familiar with evaluating linear regression models by examining the proportion of outcome variance, or R^2, that is accounted for by the model. Generalized linear models employ maximum likelihood estimation, which differs from the estimation procedures used in linear regression, necessitating a different approach for evaluating model fit. Fit is important for a Poisson model because the Poisson distribution is not very flexible, given that it has only one parameter, the mean, and the variance is required to be equal to the mean. In a generalized linear model output, the Pearson χ^2 statistic, which compares observed and expected values, $[\sum \frac{(Observed-Expected)^2}{Expected}]$, divided by the degrees of freedom (*df*) is an overall indication of how well the chosen distribution fits the data. The deviance, which is based on the likelihood, divided by degrees of freedom is another measure of distribution fit. The requirement that the mean equal the variance is an assumption that, if violated, results in improper estimates of the standard errors. A good fit is indicated by goodness-of-fit values in the neighborhood of 1.0. Overly small values reflect underdispersion, while excessively large values reflect overdispersion. The problem in Poisson model fit is typically overdispersion, indicating that the variance is larger than the mean, and there can be several reasons for this, including (a) missing interaction terms; (b) outliers in the data; and (c) unexplained heterogeneity, indicating hidden subgroups of individuals in the data that have different rates of events. In our example, there are several couples with large numbers of physical assaults, which may distort the variance. Overdispersion may lead to underestimates of the standard errors, which may falsely reduce the width of the confidence intervals, thus increasing the rate of type I error.

Aside from adding effect modifiers to the model or removing erroneous outliers, there are two main approaches to handling overdispersion when using SAS (Pedan, 2001). The first method, using the DSCALE or PSCALE option in the SAS model statement, simply applies a correction to the variance calculation. Using the scaling options, estimates are left unchanged while the standard errors are inflated or deflated to accommodate overdispersion or underdispersion, respectively.[2] (If using SPSS, the syntax we have provided above also corrects for underestimation of the standard errors, although the approach is somewhat different from the SAS DSCALE and PSCALE options.)

The second approach is to change the distribution from Poisson to a negative binomial distribution. The negative binomial distribution can be considered a Poisson distribution where the variance is allowed to be greater than the mean (Pedan, 2001). To model a negative binomial distribution using SAS, specify the DIST = NB and LINK = LOG options in the PROC GENMOD model statement.

Interpretation of Results

In a manner similar to logistic regression, Poisson models are estimated on a logarithmic scale; therefore, obtaining an interpretable result requires that effect estimates be exponentiated. Instead of an odds ratio, the parameter estimates provide a rate ratio, also referred to as an *incidence density ratio* when the denominator is time. As with other regression models, when there are multiple predictors in the model, results are interpreted as being adjusted for, or conditional on, the other predictors. In performing a multivariate analysis, one should consider Poisson regression as a large sample procedure. The rule of thumb used for logistic regression, calling for 5 to 10 events per predictor variable in the model, is most likely appropriate for Poisson regression as well; however, more information can be found in Signorini (1991).

AN EXAMPLE OF RESULTS USING DOMESTIC VIOLENCE DATA

To illustrate differences in the utility of linear regression and Poisson regression for count outcomes, we ran the following three models using our example data: (a) linear regression, (b) Poisson regression without scale correction for overdispersion, and (c) Poisson regression corrected for overdispersion, as presented in Figure 2, Panels A, B, and C. Each of the three models regressed P_AGGRESSION, the CTS2 aggression outcome variable ($M = 11.46$; $SD = 25.79$; $n = 114$), on BMIDS, the alcohol use predictor variable ($M = 46.65$; $SD = 25.93$; $n = 114$). Shown in Panel A is the linear regression model syntax and output. The p-values for the intercept and predictor, BMIDS, are shown in the column labeled "Pr > | t |." We interpret the output of this analysis as indicating that there is an estimated increase of 0.15 acts of physical aggression for each unit increase in BMIDS score, a result that is not significant at an $\alpha = .05$ level. As a one unit change in BMIDS is not particularly relevant, a common practice with continuous variables is to use the standard deviation as the unit of change. The standard deviation for BMIDS is 25.93, indicating an estimated increase of 3.89 acts of physical aggression (0.15 * 25.93) for each standard deviation unit increase in BMIDS. The distribution of the residuals for this linear regression analysis further indicated that the variance of the residuals is not normally distributed, a violation of one of the assumptions of linear regression.

Results of our two Poisson regression models are presented in Figure 2, Panel B and Panel C. For each model, p-values for the predictor, BMIDS, are shown in the column labeled "Pr > ChiSq." Our first Poisson regression analysis output, shown in Figure 2, Panel B, is not usable as is because in examining the scaled deviance/*df* we observe a value of 28.74, indicating significant overdispersion. Using the DSCALE option in our second Poisson analysis, as shown in Figure 2, Panel C, results in a corrected scaled deviance/*df* that is equal to 1.00. Comparison of these two Poisson regressions (Figure 2, Panel B and Panel C) highlights that although the resulting estimates of the effects of the alcohol use predictor BMIDS are the same (0.0127), the standard error of estimate for the Poisson analysis that corrected for overdispersion (Figure 2, Panel C) is approximately 5 times larger (e.g., overdispersed = 0.0011 vs. corrected for overdispersion = 0.0058). As shown in Panel B of Figure 2, failure to correct for overdispersion results in type I error, or rejecting the null hypothesis when it is true. Interpretation of the parameter estimate requires exponentiating it, therefore, $e^{0.0127} = 1.013$, (where $e = 2.718\ldots$). This represents the effects of a one unit change in BMIDS and because a single unit change is not very informative, again, we use the standard deviation as the unit of change, which yields $e^{0.0127*25.93} = 1.39$ as our incidence density ratio. This implies that the risk of a physical assault is 1.4 times significantly greater with a one standard deviation unit increase in BMIDS.

CONCLUSIONS

Although this example demonstrates that significant associations were obtained with Poisson regression and not with linear regression, the main point of this comparison is that linear regression is not robust to violations of its assumptions. Count data, generally speaking, will follow a Poisson distribution; therefore, it is preferable to use Poisson regression, which corresponds to the distribution of the outcome data. This may be particularly important in trauma studies of behavioral outcomes, such as aggressive acts or substance abuse, since variables representing these constructs are highly characterized by low or zero scores. Regarding practical

[2] As explained in greater detail by Pedan (2001), in a Poisson model, overdispersion may be handled by including a dispersion parameter in the relationship between the mean and the variance. McCullagh and Nelder (1989) proposed estimating the dispersion parameter as the ratio of the Pearson χ^2 statistic or the deviance to the degrees of freedom. This approach can be applied in SAS by including "PSCALE" or "DSCALE" in the model statement.

```
Panel A. Linear Regression Model

PROC REG DATA = trauma2;
MODEL P_AGGRESSION = BMIDS;
RUN;
                         Parameter Estimates
                      Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|

Intercept    1        4.67930        4.95754       0.94      0.3473
BMIDS        1        0.14545        0.09298       1.56      0.1206

Panel B. Poisson Regression Model, With Overdispersion

PROC GENMOD DATA = trauma2;
MODEL P_AGGRESSION = BMIDS / DIST=POISSON LINK = LOG OFFSET = LTIME;
RUN;

             Criteria For Assessing Goodness Of Fit
  Criterion                 DF           Value        Value/DF

  Deviance                 112       3219.3804         28.7445
  Scaled Deviance          112       3219.3804         28.7445
  Pearson Chi-Square       112       6666.3543         59.5210
  Scaled Pearson X2        112       6666.3543         59.5210
  Log Likelihood                     1950.9578

                Analysis Of Parameter Estimates
                         Standard      Wald 95%
Parameter  DF  Estimate     Error   Confidence Limits  Pr > ChiSq

Intercept   1    0.0037    0.0654   -0.1245    0.1319   0.9551
BMIDS       1    0.0127    0.0011    0.0105    0.0148   <.0001
Scale       0    1.0000    0.0000    1.0000    1.0000

Panel C. Poisson Regression Model, Corrected for Overdispersion

PROC GENMOD DATA=trauma2;
MODEL P_AGGRESSION = BMIDS / DIST = POISSON LINK = LOG OFFSET = LTIME DSCALE;
RUN;

             Criteria For Assessing Goodness Of Fit
  Criterion                 DF           Value        Value/DF

  Deviance                 112       3219.3804         28.7445
  Scaled Deviance          112        112.0000          1.0000
  Pearson Chi-Square       112       6666.3543         59.5210
  Scaled Pearson X2        112        231.9178          2.0707
  Log Likelihood                       67.8725

                Analysis Of Parameter Estimates
                         Standard      Wald 95%
 Parameter  DF  Estimate     Error   Confidence Limits   Pr > ChiSq

 Intercept   1    0.0037    0.3507   -0.6837    0.6911     0.9916
 BMIDS       1    0.0127    0.0058    0.0013    0.0240     0.0282
```

Figure 2. Analysis of domestic violence data using three regression methods in SAS. The outcome is number of physical aggression acts by the male partner during the prior 6 months at baseline. The predictor, BMIDS, is the baseline score on the Inventory of Drinking Situations. Panel A: Linear regression model. Panel B: Poisson regression model not corrected for overdispersion. Panel C: Poisson regression model corrected for overdispersion.

utility, Poisson regression provides a result that is readily interpretable as a rate ratio, which cannot be said when using transformed data in a linear regression, and by using PROC GENMOD, Poisson models can be easily applied to count data in trauma research.

We hope this example of application of Poisson regression will be useful to trauma researchers. As noted by Cohen et al. (2003), the Poisson approach has not been highly represented among published studies in the behavioral sciences to date. Particularly few trauma research studies have utilized Poisson regression; in fact, we are aware of only several such studies (e.g., Mills, Teesson, Ross, & Peters, 2006; Schneiderman, Braver, & Kang, 2008; Schnurr, Spiro, Aldwin, & Stukel, 1998). For those interested in further information, there are additional resources worth investigating. A SAS-oriented look at Poisson regression can be found in Stokes et al. (2001). Examples of Poisson regression for psychology research can be found in Gardner et al. (1995). Additionally, more general discussion of generalized linear models can be found in Lindsay (1997), Long (1997), or Cohen et al. (2003), as well as in the references cited below.

REFERENCES

Annis, H. M. (1982). Inventory of drinking situations. Toronto, Ontario: Addiction Research Foundation.

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

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological Bulletin, 118, 392–404.

Lindsay, J. K. (1997). Applying generalized linear models. New York: Springer-Verlag.

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

Mattson, R. E., O'Farrell, T. J., Monson, C. M., Taft, C. T., & Panuzio, J. (2008). Wives' verbal aggression predicts relapse in a treatment sample of men with substance use disorders. Manuscript submitted for publication.

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

Mills, K. L., Teesson, M., Ross, J., & Peters, L. (2006). Trauma, PTSD, and substance use disorders: Findings from the Australian National Survey of Mental Health and Well-being. American Journal of Psychiatry, 163, 651–658.

Mosteller, F., & Tukey, J. W. (1977). Data analysis and regression: A second course in statistics. Reading, MA: Addison-Wesley.

Pedan, A. (2001, April). Analysis of count data using SAS. Paper presented at the 26th International SAS Conference, Long Beach, CA.

SAS Institute. (2004). SAS/STAT 9.1 User's Guide. Cary, NC: Author.

Schneiderman, A. I., Braver, E. R., & Kang, H. K. (2008). Understanding sequelae of injury mechanisms and mild traumatic brain injury incurred during the conflicts in Iraq and Afghanistan: Persistent postconcussive symptoms and posttraumatic stress disorder. American Journal of Epidemiology, 167, 1446–1452.

Schnurr, P. P., Spiro, A., Aldwin, C. M., & Stukel, T. A. (1998). Physical symptom trajectories following trauma exposure: Longitudinal findings from the Normative Aging Study. Journal of Nervous and Mental Disease, 186, 522–528.

Signorini, D. F. (1991). Sample size for Poisson regression. Biometrika, 78, 446–450.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The Revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. Journal of Family Issues, 17, 283–316.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2001). Categorical data analysis using the SAS system (2nd ed.). Cary, NC: SAS Institute.

UCLA Academic Technology Services. (n.d.). SPSS data analysis examples: Poisson regression. Los Angeles, CA: Author. Retrieved May 28, 2008, from http://www.ats.ucla.edu/stat/spss/dae/poissonreg.htm