3-1-1996

Multiple Approaches to Absenteeism Analysis

Michael C. Sturman
Cornell University, mcs5@cornell.edu

Multiple Approaches to Absenteeism Analysis

Abstract
Absenteeism research has often been criticized for using inappropriate analysis. Characteristics of absence data, notably that it is usually truncated and skewed, violate assumptions of OLS regression; however, OLS and correlation analysis remain the dominant models of absenteeism research. This piece compares eight models that may be appropriate for analyzing absence data. Specifically, this piece discusses and uses OLS regression, OLS regression with a transformed dependent variable, the Tobit model, Poisson regression, Overdispersed Poisson regression, the Negative Binomial model, Ordinal Logistic regression, and the Ordinal Probit model. A simulation methodology is employed to determine the extent to which each model is likely to produce false positives. Simulations vary with respect to the shape of the dependent variable's distribution, sample size, and the shape of the independent variables' distributions. Actual data,based on a sample of 195 manufacturing employees, is used to illustrate how these models might be used to analyze a real data set. Results from the simulation suggest that, despite methodological expectations, OLS regression does not produce significantly more false positives than expected at various alpha levels. However, the Tobit and Poisson models are often shown to yield too many false positives. A number of other models yield less than the expected number of false positives, thus suggesting that they may serve well as conservative hypothesis tests.

Keywords
work, absenteeism, regression, OLS, research, data, model, research, positive

Disciplines
Human Resources Management

Comments
Suggested Citation
Sturman, M. C. (1996). *Multiple approaches to absenteeism analysis* (CAHRS Working Paper #96-07). Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.
http://digitalcommons.ilr.cornell.edu/cahrswp/178

CAHRS

Center for Advanced Human Resource Studies

CAHRS / Cornell University
187 Ives Hall
Ithaca, NY 14853-3901 USA
Tel. 607 255-9358
www.ilr.cornell.edu/CAHRS/

WORKING PAPER SERIES

Multiple Approaches to Absenteeism Analysis

Michael C. Sturman

Working Paper 96-07

Multiple Approaches to Absenteeism Analysis

Michael C. Sturman
Department of Human Resource Studies
School of Industrial and Labor Relations
Cornell University
393 Ives Hall
Ithaca, New York 14853-3901
Phone: (607) 255-5427
Fax: (607) 255-1836
E-mail: mcs5 c@cornell.edu

Working Paper 96-07

http://www.ilr.cornell.edu/cahrs

An earlier version of this paper was presented at the 1996 Academy of Management Meetings, Research Methods Division. Special thanks to Kim Goyette, Steve Caldwell, Tove Hammer, and John Boudreau.

Abstract

Absenteeism research has often been criticized for using inappropriate analysis. Characteristics of absence data, notably that it is usually truncated and skewed, violate assumptions of OLS regression; however, OLS and correlation analysis remain the dominant models of absenteeism research.

This piece compares eight models that may be appropriate for analyzing absence data. Specifically, this piece discusses and uses OLS regression, OLS regression with a transformed dependent variable, the Tobit model, Poisson regression, Overdispersed Poisson regression, the Negative Binomial model, Ordinal Logistic regression, and the Ordinal Probit model. A simulation methodology is employed to determine the extent to which each model is likely to produce false positives. Simulations vary with respect to the shape of the dependent variable's distribution, sample size, and the shape of the independent variables' distributions. Actual data, based on a sample of 195 manufacturing employees, is used to illustrate how these models might be used to analyze a real data set.

Results from the simulation suggest that, despite methodological expectations, OLS regression does not produce significantly more false positives than expected at various alpha levels. However, the Tobit and Poisson models are often shown to yield too many false positives. A number of other models yield less than the expected number of false positives, thus suggesting that they may serve well as conservative hypothesis tests.

A fundamental problem with scientific research is that the way we try to solve a problem affects what kind of results we see (Kuhn, 1970). This is partly reflected in the social sciences as the statistical method used to analyze data affects what kind of relationships we observe. When the assumptions of the employed statistical model are met, the observed coefficients are usually reliable and efficient (Greene, 1993), or in other words, they describe the actual relationship well. However, when these assumptions are violated, such as when using Ordinary Least Squares (OLS) regression to analyze non-normal data, resulting estimates may not be meaningful. This can result in true relationships not being discovered (i.e., Type II errors), or the mis-identification of non-existent relationships (i.e., Type I errors).

An area of human resource research that has been plagued by methodological concerns is research on absenteeism. The largest problem stems from the fact that rates of absenteeism do not follow a normal distribution (Arbous & Sichel, 1954; Baba, 1990; Hammer & Landau, 1981; Harrison & Hulin, 1989; Landy, Vasey, & Smith, 1984; Martocchio & Harrison, 1993; Mikalachki & Gandz, 1982; Rushmore & Youngblood, 1979; Watson, Driver, & Watson, 1985). This results in a skewed, truncated distribution, which contradicts the assumptions of commonly employed statistical methods, such as correlation analyses and OLS regression (Nunnally, 1978). Yet, despite fundamental methodological flaws, correlation and multiple regression dominate absence research (Baba, 1990; Martocchio & Harrison, 1993).

The methodological issues raised by the distributional characteristics of absenteeism data are quite significant. Indeed, one might even argue that prior statistical findings using correlations or OLS regression are questionable, perhaps suggesting that empirical absenteeism research should begin anew. A more conservative approach would entail evaluating the significance of these methodological problems.

Although statistical and methodological research provides alternatives to traditional analysis, the field is still largely silent regarding the implications of more typical methodologies to our understanding of absenteeism. This piece will address this point by comparing eight different analysis methods: OLS regression, OLS regression with the absenteeism variable transformed, the Tobit model, Poisson regression, Overdispersed Poisson regression, the Negative Binomial model, Ordinal Logistic regression, and Ordinal Probit regression. The models will be compared in two ways. First, a simulation will be employed to evaluate the models' propensities for Type I errors. This simulation will demonstrate the extent that these models are likely to detect statistically significant relationships when these relationships do not really exist, and will explore the sensitivity of the models to characteristics of the dependent variable's distribution, characteristics of the independent variables' distributions, and sample size. Second, using a basic data set (i.e., not a simulation), findings from the eight models will be compared. The results will help illustrate how these models might be used and interpreted in field research.

Methods of Coping with Absence-Like Data

Researchers are beginning to identify alternatives to help overcome the methodological problems associated with absenteeism research, and more generally with analyzing count data. These alternatives can be seen as either, one, changing the characteristics of the data to meet the assumptions of traditional statistical methods better, or two, using a statistical method that is more appropriate for the type of data collected.

One way to change the data involves changing the level of aggregation. Because there is a low base rate of absenteeism on any given day, aggregation allows for a wider distribution of values. This approach ostensibly produces distributions that better meet the assumptions of traditional analyses. Additionally, if we assume that there are no differing effects on absenteeism over time, then correlations at the individual level data should equal correlations of the aggregated data (Ostroff, 1993).

However, aggregation may still leave problems and create other problems that make its use undesirable (Hulin & Rousseau, 1980). For example, Harrison and Hulin (1989) show that aggregation of voluntary absence data from one month to one year exaggerates the level of skew and kurtosis, and the truncation problem is still obvious. In their example, aggregation only reduced the effect of discreteness. Thus, researchers are still faced with the methodological difficulties they started with. Additionally, aggregation may obscure relationships because of longer cause effect time gaps and may occlude the effects of some environmental variables (Harrison & Hulin, 1989). For example, if winter weather has a significant effect on an individual's ability to attend work, aggregating over the entire year will obscure the effect that winter storms have on level of absenteeism.

Another approach to studying absenteeism is to observe a related, but better behaved, dependent variable. One approach entails studying the more broadly defined construct of withdrawal behaviors, which may include absenteeism, lateness, intent to leave the organization, turnover, etc. (Hulin, 1991). This approach may yield a continuously distributed dependent variable that overcomes the problems associated with studying low base rate events (Hulin & Rousseau, 1980). A similar approach, but used to study absenteeism alone, was employed by Martocchio and judge (1994). In their piece, the authors asked subjects to read scenarios that varied with respect to hypothesized absence-related factors (presence of hobby/leisure activities, presence of community/ religious activities, if it was the beginning or end of the week, kinship responsibility, presence of pressing work, and personal illness). Subjects were then asked to indicate, using a seven-point Likert-type scale, the extent to which they would likely miss work if faced with those particular circumstances. The data collected was well suited for traditional analyses (ANOVA, OLS Regression, cluster analysis); however, the method can only be applied using an experimental methodology, and thus does not lend itself to analysis where archival data (i.e., number of absences) are used.

An alternative to traditional analysis was demonstrated by Harrison and Hulin (1989), who overcame methodological problems by using an event history model. Event history models describe the states individuals are in, the time spent in those states, and the rates of movements from state to state (Harrison & Hulin, 1989). Markov models and hazard-rate models are two examples of popular event history models. Such models overcome many of the estimation problems for absence-like data because "model estimation is built on an emerging theory called quasilikelihood, which does not require that the data conform to a specified multivariate distribution (Harrison & Hulin, 1989, pg. 315)"; however, such models entail analyzing large samples over multiple observation periods, and thus may not be applicable for samples more typical of absenteeism research (Harrison & Hulin, 1989).

Modeling Strategies

<u>OLS Regression</u>

When faced with non-normal data, we can always ignore the problem and proceed as if the data were normally distributed. This practice, although not recommended for clear statistical reasons (Johnson & Wichern, 1992), is by far the most common practice in absenteeism research as illustrated by the preponderance of OLS regression (Baba, 1990; Harrison & Hulin, 1989; Martocchio & Harrison, 1993).

One implication of this is that, because OLS regression does not account for absences being truncated at zero, it can predict negative values which are clearly meaningless. Additionally, the validity of hypothesis tests in OLS regression depends on assumptions of the variance scores that are unlikely to be met in typical count data (Gardner, Mulvey, & Shaw, 1995). As a result, sampling statistics (i.e., mean and variance) may differ significantly from the true population parameters, which could lead to a loss of power, and thus Type II errors (Hammer & Landau, 1981). A skewed distribution can also lead to heteroscedasticity, which can severely effect standard errors, and lead to Type I or Type II errors (Hammer & Landau, 1981). Because for count data, like number of absences, the residuals almost always correlate positively with the predictors, the estimated standard errors of the regression coefficients are smaller than their true value, and thus the t-values associated with the regression coefficients are inflated (Gardner et al., 1995). Thus, OLS regression seems prone to Type I errors for analysis of absenteeism data.

While these assertions are methodologically sound and have face validity, the significance of this problem has not received much empirical attention. There have been some exceptions to this. OLS regression has been compared to the Tobit model (Hammer & Landau, 1981) and the Poisson and Negative Binomial models (Cameron & Trivedi, 1986; Hausman, Hall, & Griliches, 1984). These comparisons will be discussed after the specific models have been introduced.

<u>Data Transformations and OLS Regression</u>

An alternative to ignoring the non-normality problem is to transform the data to make it "more normal looking." Transformations are nothing more than expressing the data in different units (Johnson & Wichern, 1992), and thus data transformations are commonly recommended to help satisfy statistical assumptions. For count data, such as number of absences, the square root is a recommended transformation function (Johnson & Wichern, 1992).

Unfortunately, there are a number of potential problems with transforming data, particularly for absenteeism. Most obviously, data transformations do nothing to compensate for the fact that absenteeism data are truncated at zero, and thus negative absences can still be predicted (Hammer & Landau, 1981; Harrison & Hulin, 1989). Additionally, Harrison and Hulin (1989) assert that "even transformed measures of absenteeism data are often inadmissible as dependent variables in linear regression, because linear regression assumes normality of the marginal distributions of the dependent variable (pg. 300)." The modal value of the distribution will still fall near the bottom of the scale, and although transformations may make the distribution look more normal, heteroscedasticity is still likely to be a problem. This can lead to reduced correlation coefficients, loss of power, and Type II errors (Hammer & Landau, 1981; Harrison & Hulin, 1989), or inflated t-values and Type I errors (Gardner et al., 1995). However, OLS regression after transforming skewed variables has not been compared to alternative methodologies for absence data.

<u>Tobit Model</u>

Although alternatives to OLS are rare, one technique receiving some attention in absenteeism research is the Tobit model (Baba, 1990; Hammer & Landau, 1981). The Tobit model (Tobin, 1958) is a regression model designed to handle truncated data, where the truncated value occurs with a high probability and the variable is continuously distributed beyond that point (Baba, 1990; Greene, 1993). Tobit models are espoused to provide more consistent, reliable, and less biased estimates than the OLS model (Baba, 1990; Leigh, 1985; Maddala,1983).

Some research has suggested that the Tobit model is more sensitive than OLS regression (Baba, 1990; Hammer & Landau, 1981). While this assertion has received some support in the form of more significant coefficients resulting from Tobit models than OLS models (Baba, 1990), this does not necessarily prove the value of the Tobit model. If we assume that the new significant values from the Tobit model are not Type I errors (or in other words, the lack of significance from the OLS regression are Type II errors), this would imply that significant findings from previous research relying on OLS regression are still valuable. Indeed, this seems to be what researchers espousing the merits of the Tobit model are implying: they suggest that the Tobit model be used as a check on an OLS solution (Baba, 1990; Hammer & Landau, 1981). Nonetheless, the need still exists for further comparisons of the Tobit model to OLS regression (Baba, 1990). The Tobit model should also be compared to OLS with transformed data. Additionally, because the Tobit model assumes a continuous dependent variable, and because absences data are discrete, the Tobit model should be compared to count models.

<u>Poisson Regression</u>

For data where the dependent variable is a discrete count, Poisson regression is a natural model choice (Cameron & Trivedi, 1986, 1990; Gurmu, 1991; Hausman et al., 1984; Lee, 1986). Poisson models are particularly attractive for modeling count data because the model has been extended into a regression framework (Lee, 1986), it has a simple structure, and it can be easily estimated (Greene, 1993; Lee, 1986).

However, this simplicity is the result of some limiting assumptions, violations of which may have significant affects on the reliability and efficiency of the model coefficients. The most significant criticism of the Poisson model is of its assumption that the variance of the dependent variable equals its mean (Cameron & Trivedi, 1986; Greene, 1993; Lee, 1986). Poisson regression also assumes that each occurrence is independent of the number of previous occurrences, and the expected number of occurrences is identical for every member of the sample.

Research has addressed some of these limitations by developing tests for overdispersion (e.g., Cameron & Trivedi, 1986, 1990; Gurmu, 1991; Gurmu & Trivedi, 1992; Lee, 1986). Overdispersion occurs if the distribution's variance is greater than the distribution's mean (Greene, 1993). Overdispersion causes the estimates of standard errors to be less than their true value, which leads to inflated tcoefficients and Type I errors.

The distribution of absences has not been compared to the Poisson distribution in terms of overdispersion. Further, overdispersion will depend on how the data is aggregated (e.g., by the month, year, etc.). One approach for dealing with overdispersed data is to use a less constrained model (Cameron & Trivedi, 1986; Gurmu, 1991; Lee, 1986), such as the Negative Binomial model. Another approach is to correct the t-values based on an estimate of the dispersion. Despite its limiting assumptions and the availability of alternatives, the implications of Poisson regression for absenteeism research should be considered.

Overdispersed Poisson Regression Model

If overdispersion exists, one method of correcting for its implications, described by Gardner et al. (1995), is called Overdispersed Poisson Regression. This technique entails estimating a dispersion parameter and using it to modify the t-tests resulting from a Poisson regression.

The overdispersion term is a function of the squared deviation from its expected value (see Gardner et al., 1995, pg. 397):

$$\varnothing = (N - J)^{-1} \sum \{y_l - \mu\,[X_l,d_l])^2 / \ \mu\,[X_l,d_l]\ \}$$

N = Number of cases

J = Number of independent variable

y^l = Observed value

$\mu\,[X_l,d_l]$ = Predicted value

Each squared deviation is then divided by that score's variance assuming that the standard Poisson model were true (Gardner et al., 1995). If the variance equals the mean, then as the sample size approaches infinity, the deviation score will equal one. Because the $\mu\,[X_l,d_l]$ stimates are chosen to fit the specific sample of y^l, the deviation score sum needs to be adjusted so the dispersion term will still approximate one if the assumptions of the Poisson model are met with less than infinite sample size (Gardner et al., 1995).

The value of $\varnothing$can be tested to see if the overdispersion is statistically significant (Gardner et al., 1995), but there are also many other methods for testing overdispersion (e.g., Cameron & Trivedi, 1990). Of more immediate use, the value of $\varnothing$can be used to modify the results from a Poisson regression. If significant overdispersion exists but $\varnothing$ is assumed to equal one, then the estimated variance of the regression coefficients will be smaller than their true values (Gardner et al., 1995). This will result in inflated t-tests for the regression coefficients. The Poisson regression results can be corrected by multiplying the Cov (B) by $\varnothing$; or, the t-tests computed by the Poisson regression can be divided by the square root of ∅ (Gardner et al.,1995).

Use of the Overdispersed Poisson regression model, as demonstrated by Gardner et al. (1995) resulted in t-test scores that were dramatically lower than those generated by a Poisson model. The results of the Overdispersed Poisson regression model were also similar to those from a Negative Binomial regression model. Yet, it is still unclear how the Overdispersed Poisson regression model would compare to the other models described in this paper.

Negative Binomial Model

Although the Overdispersed Poisson regression model addresses the assumption of Poisson regression that the mean equals the variance, other research on absenteeism calls into question the other assumptions of the model for analyzing absence data. Research has shown that past absences are one of the best predictors of future absences (Ivancevich, 1985; Morgan & Herman, 1976; Waters & Roach, 1979), thus calling into question the independence assumption of Poisson models. Additionally, theoretical models of absenteeism (e.g., Blau & Boal, 1987; Gibson, 1966; Nicholson, 1977; Rhodes & Steers, 1990; Steers & Rhodes, 1978) commonly suggest that absenteeism is a function of the construct ability to attend work, which includes illness and accidents, family responsibilities, and transportation problems. These models suggest that individual characteristics cause (or at least correlate with) absenteeism, thus suggesting that the expected number of absences for individuals will differ. When the mean level of absences is expected to differ across cases, the Multiple Approaches 12 Negative Binomial model may be more appropriate (Gardner et al., 1995).

The Negative Binomial model is one of the more general count models (Cameron & Trivedi, 1986; Gurmu, 1991; Gurmu & Trivedi, 1992; Lee, 1986). In fact, the Poisson model is a special case of the Negative Binomial model (Cameron & Trivedi, 1986). Negative Binomial models can take a number of forms. Commonly, they are categorized through one of two specifications of the variance of the dependent variable: one, $Var(y) = (1 + a)E(y)$, or two, $Var(y) = E(y)*(1 + a)E(y)$, where a is positive. The former case implies a constant variance-to-mean ratio; the latter case implies a variance-to-mean ratio that is linear (Cameron & Trivedi, 1986). Clearly, there are even more possibilities; however, this study will focus on the first case, in part because it is simpler, but primarily because this study has a practical focus and only the former model was included within a commonly available statistics package (e.g., Greene, 1992). As noted earlier, there is a dearth of literature comparing the effects of different types of modeling methods. However, there are also some exceptions to this for the Negative Binomial model.

Cameron and Trivedi (1986) compare the methodological implications and results of analyzing count data for a number of models, including OLS, Poisson, and Negative Binomial models. Out of 12 independent variables at an alpha level of .05, the OLS model revealed four significance coefficients. The Poisson model revealed the same four significant coefficients plus five others; the Negative Binomial model revealed the same four as the OLS model, two significant coefficients also found by the Poisson model, and one significant coefficient that was not detected by either the OLS or Poisson regression models. The authors conclude by recommending a sequential modeling strategy in which one begins with the basic Poisson model and proceeds to increasingly flexible and data-coherent models (Cameron & Trivedi, 1986).

A similar illustration can be found is an article that focused on developing and adapting models of counts for panel data (Hausman et al., 1984). Although comparing model types was not the focus of the study, the piece did show results for OLS, Poisson, and Negative Binomial models. In one instance, the OLS model showed two of seven coefficients to be significant. The Poisson model showed five of the seven to be significant, and the Negative Binomial model showed three of the seven to be significant. In a second instance, the OLS model showed one of two coefficients to be significant, while both the Poisson and Negative Binomial models showed both coefficients to be significant. Although Hausman et al. (1984) do not specifically discuss the significance of these differences, they do point out that the Negative Binomial model better represents the data as illustrated by the log-likelihood values.

Both of these studies are useful demonstrations of alternative modeling strategies. Unfortunately, they do not shed much light on the meaning of significant coefficients or the implications of the various approaches. A more explicit comparison of modeling techniques to demonstrate the implications of various models for absenteeism research is needed.

Ordinal Logit and Ordinal Probit

Another way to model absence data is to take advantage of its ordered nature. The Ordered Logit and Probit models have come into fairly wide use as a means for analyzing discrete, ordered data (Zavoina & McElvey, 1975).

The models are built around a regression framework, and thus are relatively easy to estimate. Additionally, the Binomial Probit and Binomial Logit models can be seen as special cases of their corresponding ordinal models. The differences between the Ordinal Logit and Probit models are similar to the differences between their binomial counterparts. The underlying distributions of both models are similar in the middle of the distribution, but the logit model is considerably heavier at the tails (Greene, 1993). While some research has discussed justification of one model over another (Amemiya, 1981), the issue remains unresolved and does not seem to make much of a difference (Greene, 1993).

A disadvantage of these models is that interpretation of their coefficients is highly complex (Greene, 1993). Further, although absence data is ordered, ordinal analysis may not be meaningful. Specifically, the difference between four absences and three absences is conceptually the same as the difference between six absences and five absences. Additionally, ordinal models require that there are instances of each level. So, if modeling yearly absences, which may range from zero to 40, if no one was absent 25 times, the ordinal model cannot be estimated. This can be corrected by either truncating the distribution, or renaming the instances of absences above 25 to be one less. The first remedy opens an entirely different issue about the effect of truncation on models of counts, an issue that will not be addressed by this paper. The second remedy will also work, but would make interpretation of the coefficients even more confusing, and would make predictions of future levels of absences confusing at least, and perhaps impossible. In sum, ordinal models may not always be applicable to absence-like data; one the other hand, it may be appropriate in certain circumstances, and at least evaluation of the methodologies seems warranted. The author found no comparisons of ordinal models to other models for count data for more than the bivariate case.

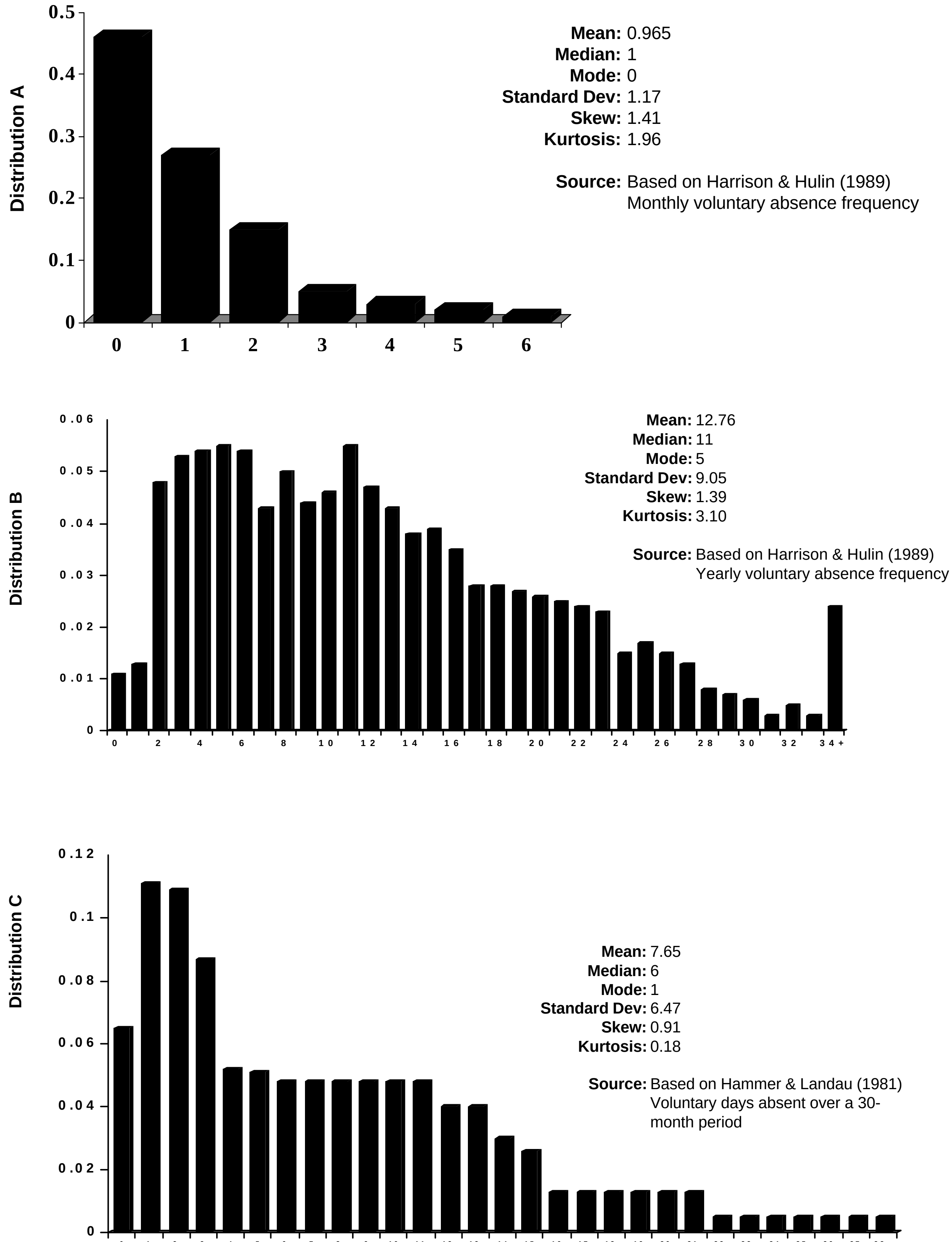
Simulation Tests

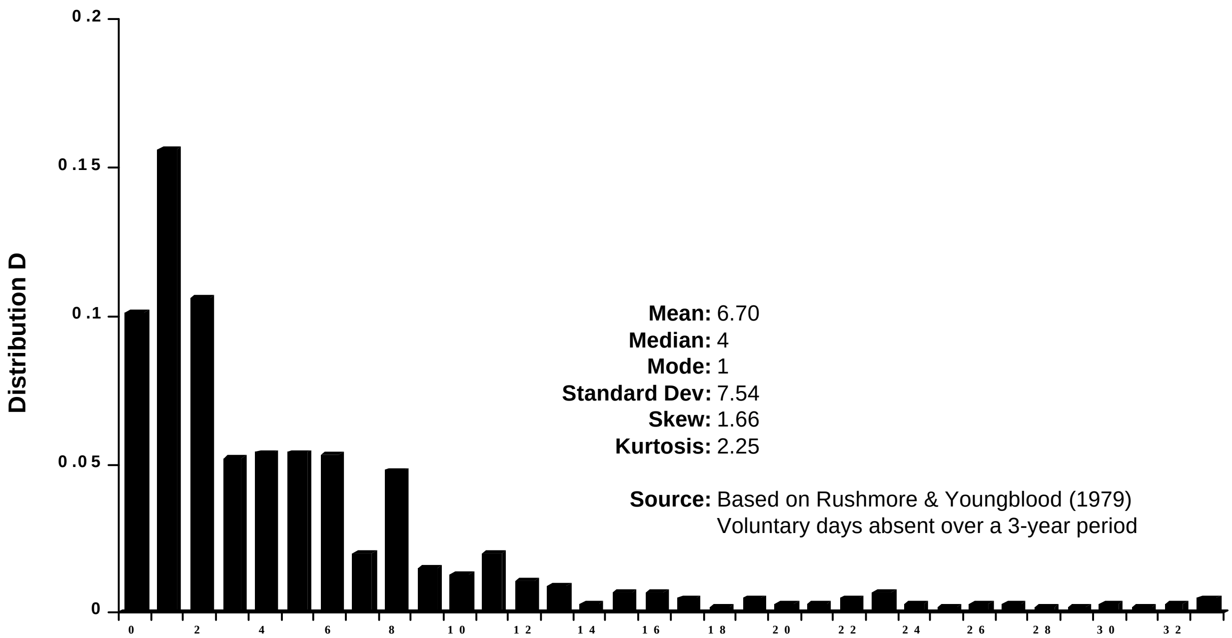
The studies that have compared various methodologies have all revealed that different models vary somewhat in terms of what coefficients are identified as significant (Baba, 1990; Cameron & Trivedi, 1986; Hausman et al., 1984). However, simply the existence of more significant coefficients does not necessary mean a model is better. Indeed, this discrepancy harks back to the issue of Type I versus Type II errors. For example, in a comparison of two models, one of which shows one more significant coefficient, either one model is exhibiting a Type I error, or the other model is exhibiting a Type II error. The stress in social science research has generally been on minimizing the chance of Type I errors. Therefore, this piece will concentrate on evaluating the extent to which these various models are prone to this sort of error. In other words, this paper will compare how likely these models are to yield false positives.

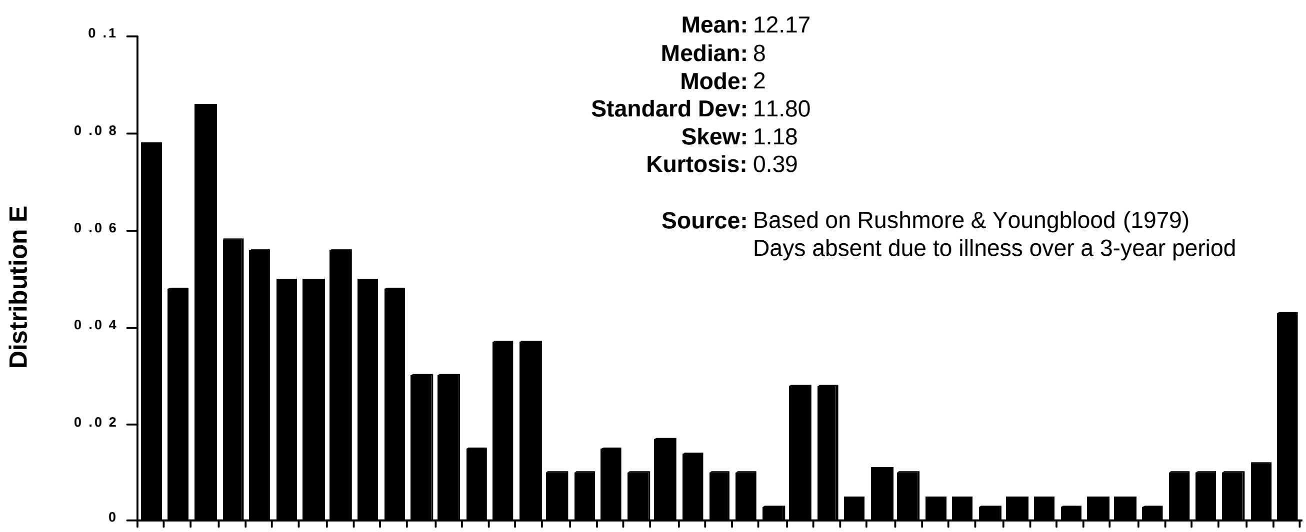
To compare the sensitivity of the above models (OLS, transformed dependent variable OLS, Tobit, Poisson, Overdispersed Poisson, Negative Binomial, Ordered Logit, and Ordered Probit), a simulation will be employed. Specifically, simulation will be used to count the number of times that each model incorrectly identifies a significant relationship.

The dependent variable will be a measure of absences. Because the way the dependent variable is generated will affect the results, we will not rely on any common mathematical distribution (e.g., a chi-square, squared normal, etc.) to generate the variable. Rather, the variable will be generated to resemble a distribution that has been reported in literature on absenteeism (Hammer & Landau, 1981; Harrison & Hulin, 1989; Rushmore & Youngblood, 1979). The five distributions that will be used include a variety of aggregations (monthly to three years) and have varying degrees of skewness and kurtosis. Additionally, because these distributions are based on actual absenteeism data, the simulations will be realistic and we avoid the problem of developing a distribution of absences that has an a priori relationship with one of the models. Because the distributions used in this paper are based on estimates from published figures rather than from the original data, the distributions in this piece may not exactly match those upon which they are based. Further, the distributions may have been changed slightly to facilitate simulation, such as truncating the distribution after the probability of a individual levels of absences dropped below one percent. The distributions used in this paper and their summary statistics are shown in Figure 1.

Figure 1. Distribution of absenteeism variables and summary statistics.




Distribution D
0.2
0.15
0.1
0.05
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32
Mean: 6.70
Median: 4
Mode: 1
Standard Dev: 7.54
Skew: 1.66
Kurtosis: 2.25
Source: Based on Rushmore & Youngblood (1979)
Voluntary days absent over a 3-year period


Distribution E
0.1
0.08
0.06
0.04
0.02
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42+
Mean: 12.17
Median: 8
Mode: 2
Standard Dev: 11.80
Skew: 1.18
Kurtosis: 0.39
Source: Based on Rushmore & Youngblood (1979)
Days absent due to illness over a 3-year period

The simulation will estimate the effects of model type, distribution of the dependent variable, sample size, and distribution of the independent variables. The eight models and five distributions have been described above. To determine if sample size impacts the sensitivity of these models to errors, simulations will be run with samples of 100 and 1000. Because past absences are frequently used to predict future absences (e.g., Ivancevich, 1985; Clegg, 1983), some simulations will be run with the independent variable following the same distribution as the dependent variable. This will allow the simulation to determine if the distribution of the independent variable affects the likelihood of Type I errors.

The simulation variables will be completely crossed, which permits the assessment of the independent effects of each factor on the likelihood of Type I errors. Crossing the factors yields 160 simulations scenarios (8 x 5 x 2 x 2). For each simulation, the dependent variable will be generated from one of the absence distributions shown in Figure 1. Each model will have 10 independent variables drawn either from a normal distribution (mean = 0; SD = 1) or the same distribution as the dependent variable. Either 100 or 1000 cases will be randomly generated, with each variable being generated independently. Thus, there is no a priori relationship between the random independent variables and the dependent variable. For each scenario, 50 simulations will be performed. For each set of generated data, each of the eight models described above will be used to determine the relationship between the independent variables and the dependent variable. Any significant findings are spurious relationships, and thus Type I errors.

The number of Type I errors will be recorded at the $p<.10$, $p<.05$, and $p<.01$ levels. If significantly more than the expected number of Type I errors occur, such as more than 10 percent at alpha equals .1.0, then we can conclude that the model is prone to Type I errors. Although not a direct test of Type II errors, if significantly less than the expected number of coefficients are statistically significant, then the t-tests may be too conservative, and thus are prone to Type II errors.

The proportion of Type I errors by model are shown in Table 1. T-tests reveal that, for a number of models, the number of Type I errors significantly differs from the expected level. Most obviously, the Poisson model incorrectly identifies far more significant relationships than expected by the alpha level. However, the Tobit model also shows significantly more false positives than would be expected by chance. One the other hand, some models identify significantly fewer false positives than the alpha levels would suggest. The Ordered Probit models has fewer Type I errors at alpha=.10, and the Ordered Probit model has fewer Type I errors at the alpha=.10 and alpha=.05 levels. The Negative Binomial model has fewer false positives at all three alpha levels.

Table 1

Type I Errors by Model

Model	N	Mean Number of Type I Errors at p < .10	Mean Number of Type I Errors at p < .05	Mean Number of Type I Errors at p < .01
OLS	1000	0.99 (0.96)	0.47 (0.67)	0.10 (0.33)
Transformed OLS	1000	0.96 (0.95)	0.50 (0.70)	0.10 (0.32)
Tobit	1000	1.12* (1.06)	0.57* (0.78)	0.14* (0.39)
Poisson	1000	4.65** (2.15)	3.91 ** (2.09)	2.70** (1.86)
Overdispersed Poisson	1000	1.04 (1.01)	0.54 (0.73)	0.12 (0.37)
Negative Binomial	1000	0.71** (0.87)	0.33** (0.62)	0.05** (0.25)
Ordered Logit	571	0.88* (0.91)	0.45 (0.70)	0.10 (0.32)
Ordered Probit	571	0.87* (0.92)	0.43* (0.68)	0.08 (0.29)
Expected		**1.00**	**0.50**	**0.10**

* p<.05;**p<.0001

Each simulation contained 10 independent variables. Significance indicates deviation from the expected number of Type I errors. 1000 simulations were conducted for each model. Because the Ordered Logit and Ordered Probit required that there be no missing values in the middle of a distribution, the models were incalculable in a number of instances, particularly when the simulation was conducted with the smaller sample size.

These results suggest that Poisson regression is clearly inappropriate for analyzing absence data, unless the model is corrected for overdispersion. The Tobit model is also prone to Type I errors. Although these simulations did not directly test Type II errors, the Ordered Logit, Ordered Probit, and Negative Binomial models appear to be more conservative than the chosen alpha levels imply. This may suggest that the models might not identify true relationships, but further research is needed to clarify this issue. Nonetheless, it appears that significant results, such as from the Negative Binomial model, are less likely to be Type I errors than significant results from other models. Of note in these results is that OLS regression yields the number of Type I errors expected by chance. This implies that previous results using OLS regression are valid, or at least are not Type I errors.

MANOVA was used to determine if the characteristics of the analysis affect the number of Type I errors. The dependent variables were the number of false positives at all three alpha levels. Simulation characteristics and first order interactions were the independent variables. All the independent variables were treated as categorical. Results of the MANOVA are shown in Table 2.

TABLE 2

Multivariate Analysis of Variance Results

Source	DF	F	Wilks' Lambda
Main Effects			
Analysis Model	21, 20316	581.26	0.26**
Distribution of Independent Variable	12,18719	41.36	0.93**
Sample Size	3, 7075	3.58	1.00*
Distribution of Dependent Variable	3, 7075	4.42	1.00**
Interactions			
Analysis Model x Distribution of Ind. Var	84, 21166	47.73	0.60**
Analysis Model x Sample Size	1, 20316	1.23	1.00
Analysis Model x Distribution of Dep. Var.	21, 20316	0.81	1.00
Distribution of Ind. Var. x Sample Size	12,18719	3.04	1.00**
Distribution of Ind. Var. x Distribution of Dep. Var	12,18719	2.39	1.00**
Sample Size x Distribution of Dep. Var	3,7075	4.83	1.00**
Total	192, 21214	77.28	0.20**

* p<.05;**p<.01

Results suggest that all the main effects are significant, as are a number of interactions. The model type, distribution of the dependent variable, sample size, and distribution of the independent variables all affect the likelihood of Type I errors. Additionally, this likelihood depends on the interactions between the distribution of the dependent variable and the model type, the sample size, and the distribution of the independent variables. The effect of sample size also depends on the distribution of the independent variables. Based on these results, it is possible to estimate which models produce too many or too few Type I errors. These estimates, grouped by simulation characteristics, are shown in Table 3.

Table 3
Model Performance Under Simulation Scenarios

Distribution of Ind. Variable	Sample Size	Distribution of Dep. Var.	Models Performing Significantly More Conservatively Than Expected	Models Performing Appropriately	Models Performing Significantly Less Conservatively Than Expected
A	100	Normal		Logit, Neg-Bin, OLS, T-OLS, Probit, Ov-Poisson	Tobit, Poisson
A	100	Absent-like		Logit, Neg-Bin, OLS, T-OLS, Probit, Ov-Poisson	Tobit, Poisson
A	1000	Normal		Logit, Neg-Bin, OLS, T-OLS, Probit, Ov-Poisson, Tobit	Poisson
A	1000	Absent-like		Logit, Neg-Bin, OLS, T-OLS, Probit Ov-Poisson Tobit	Poisson
B	100	Normal	Neg-Bin	OLS, T-OLS, Ov-Poisson	Tobit, Poisson
B	100	Absent-like		Neg-Bin, T-OLS,	OLS, Ov-Poisson Tobit, Poisson,
B	1000	Normal	Logit, Neg-Bin, Ov-Poisson, OLS, T-OLS, Probit, Tobit		Poisson
B	1000	Absent-like	Neg-Bin	Logit, OLS, T-OLS, Probit Ov-Poisson Tobit	Poisson
C	100	Normal	Neg-Bin	OLS, T-OLS, Ov-Poisson	Tobit, Poisson
C	100	Absent-like	Neg-Bin	OLS, T-OLS, Ov-Poisson, Tobit	Poisson
C	1000	Normal	Neg-Bin	Logit, OLS, T-OLS, Probit, Ov-Poisson, Tobit	Poisson
C	1000	Absent-like	Neg-Bin	Logit, OLS, T-OLS, Probit Ov-Poisson Tobit	Poisson
D	100	Normal		Neg-Bin, OLS, T-OLS, Ov-Poisson	Tobit, Poisson
D	100	Absent-like		Ne -Bin T-OLS	OLS, Ov-Poisson, Poisson, Tobit
D	1000	Normal	Neg-Bin, T-OLS, Logit, Probit	OLS, Ov-Poisson, Tobit	Poisson
D	1000	Absent-like		Logit, Neg-Bin, OLS, T-OLS, Probit Ov-Poisson Tobit	Poisson
E	100	Normal	Neg-Bin, OLS	T-OLS, Ov-Poisson, Tobit	Poisson
E	100	Absent-like	Neg-Bin	OLS, T-OLS, Ov-Poisson, Tobit	Poisson
E	1000	Normal	Neg-Bin	Logit, OLS, T-OLS, Probit, Ov-Poisson, Tobit	Poisson
E	1000	Absent-like	Neg-Bin	Logit, OLS, T-OLS, Probit Ov-Poisson Tobit	Poisson

Note: Models classified as performing more conservatively than expected had significantly ($p < .05$) less Type I errors than expected by chance. Models classified as performing less conservatively than expected had significantly ($p < .05$) more Type I errors than expected by chance.

Although the characteristics of the simulation do not vary greatly (e.g., only two sample sizes are considered), it is possible to treat some of these characteristics as continuous numerical variables (treat sample size as continuous and substitute the skew and kurtosis as continuous numerical independent variables instead of the categorical variable of distribution type) and use regression analysis to predict how skew, kurtosis, and sample size influence the number of Type I errors. These formulae can then be used to estimate the likelihood of Type I errors for any distribution or sample size. Because of the size of the formulae, they are not reported here; they are, however, available from the author.

Simulations were not run for Type II errors. This is because a standard simulation methodology would produce biased results. The way a model was specified, even the way an error or noise term would be added to the model, would bias the simulation in favor of one type of model over another. Indeed, if the "true" relationship underlying number of absences was known, research such as this piece would be unnecessary. Instead, this paper will apply these models to actual data to demonstrate how typical absence research would be affected by model choice.

Interpreting Actual Data

Data are from blue collar workers at a midwestern company. Two absence measures were collected: number of excused absences, and number of unexcused absences. For an absence to be excused, the worker needs to obtain his or her supervisor's approval. Thus, calling in sick would generally be considered an excused absence; however, simply not showing up to work counted as an unexcused absence. Absence data was aggregated over a year. The independent variables include the number of days of each type of absence from the previous year, age, tenure, sex, marital status, and number of children. A total of 198 employees worked at the firm for all of 1991 and 1992, and of these, there was complete data on 195 subjects.

Note that the intent of this piece is not to provide a test of a theoretical model with absences; rather, this paper will simply illustrate the different effects of various models. The models will be realistic, though, in that they will resemble what a researcher might generate. Specifically, the models will look for significant effects of the aforementioned variables for predicting each kind of absence and will also test the significance of the interaction of lagged absences and sex (i.e., to determine if the relationship between past absences and future absences differs by sex). Summaries of the variables used in this study are shown in Table 4.

TABLE 4

Summary of Actual Data

	Mean	Std	1	2	3	4	5	6	7	8
1. Excused 1992 Absences	2.28	3.26	1.00							
2. Unexcused 1992 Absences	0.56	0.87	.35	1.00						
3. Excused 1991 Absences	1.93	2.58	.48	.26	1.00					
4. Unexcused 1991 Absences	0.39	0.74	.30	.31	.21	1.00				
5. Age	35.47	8.85	.07	-.13	.07	-.03	1.00			
6. Tenure	8.68	5.94	-.17	-.27	-.13	-.04	.45	1.00		
7. Sex (1 = Female; 0 = Male)	0.34	0.47	.32	.16	.18	.05	.03-	.29	1.00	
8. Marital Status (1 = Married; 0 = Single)	0.82	0.39	-.06	-.09	.00	-.05	.28	.14	-.08	1.00
9. Number of Children	1.84	1.57	-.07	-.14	-.03	-.05	.12	.06	.01	.43

Note: N =198. Coefficients greater than .14 are significant at $p < .05$.

1. The skew of Excused 1992 Absences is 1.99. Its kurtosis is 3.99.
2. The skew of Unexcused 1992 Absences is 1.69. Its kurtosis is 2.79.
3. The skew of Excused 1991 Absences is 1.63. Its kurtosis is 2.51.
4. The skew of Unexcused 1991 Absences is 2.08. Its kurtosis is 3.92.

Results from the eight models are shown in Tables 5 and 6. Of course, the "true" relationships are unknown; rather, the situation is similar to what any researcher would face when presented with statistical results. The comparison of models, though, helps illustrate the different effects of the models; the differences between the models highlight the difficulty associated with absenteeism research.

Table 5

Model Outputs for Actual Data: Dependent Variable is Number of Excused Absences in 1992

	Model							
Variable	OLS	Transformed Variable/OLS	Tobit	Poisson	Overdispersed Poisson	Negative Binomial	Ordinal Logit	Ordinal Probit
'91 Excused Absences	0.27 (0.10)*	0.38 (0.088)**	0.54 (0.16)**	0.13 (0.024)**	0.13 (0.043)**	0.17 (0.065)**	0.26 (0.078)**	0.15 (0.047)**
Age	0.033 (0.026)	0.0039 (0.0088)	0.018 (0.041)	0.012 (0.0058)*	0.012 (.010)	0.0061 (0.012)	-0.0027 (0.018)	0.0045 (0.011)
Tenure	-0.065 (.039)+	-0.017 (0.013)	-0.097 (0.064)**	-0.035 (0.010)**	-0.035 (0.019)*	-0.038 (0.019)*	-0.033 (0.028)	-0.023 (0.017)
Sex	0.19 (0.56)	0.24 (0.22)	1.02 (0.91)	0.44 (0.15)**	0.44 (.27)	0.51 (0.30)+	0.47 (0.41)	0.29 (0.26)
Married	-0.57 (0.58)	-0.012 (0.20)	-0.48 (0.92)	-0.24 (0.14)+	-0.24 (.24)	-0.093 (0.29)	-0.11 (0.40)	-0.52 (0.24)
# Children	-0.056 (0.14)	-0.011 (0.047)	-0.13 (0.22)	-0.015 (0.031)+	-0.015 (0.056)	-0.0043 (0.080)	-0.022 (0.097)	-0.014 (0.058)
'91 Absences *Sex	0.58 (0.16)**	0.22 (0.14)	0.46 (0.25)+	0.040 (0.031)	0.040 (.055)	0.0098 (0.11)	0.20 (0.12)+	0.11 (0.073)
R^2	0.34	0.29	-	-	-	-	-	-
Pseudo R^2	-	-	0.32	0.35	0.35	0.31	0.31	0.31
Log Likelihood	-465.34	-476.81	-359.89	-439.32	-439.32	-362.15	-347.12	-345.50

Note: N =195
\+ p < .10
* p<.05
** p<.01

1. Sigma equals 3.92, with a standard error of 0.29, and is significant at p < .01.
2. Overdispersion, ∅, equaled 3.18, and is significant at p < .01.
3. Alpha equals 1.15, with a standard error of 0.22, and is significant at p < .01.

Table 6

Model Outputs for Actual Data: Dependent Variable is Number of Unexcused Absences in 1992

	Model							
Variable	OLS	Transformed Variable/OLS	Tobit	Poisson	Overdispersed Poisson	Negative Binomial	Ordinal Logit	Ordinal Probit
'91 Unexcused Absences	0.41 (0.10)**	0.31 (0.096)**	0.83 (0.24)**	0.57 (0.12)**	0.57 (.13)**	0.57 (0.13)**	0.88 (0.24)**	0.53 (0.14)**
Age	-0.001 (0.077)	-0.0020 (0.054)	-0.0047 (0.019)	-0.000017 (0.012)	-0.000017 (.013)	-0.000021 (0.014)	-0.0077 (0.021)	-0.0018 (0.012)
Tenure	-0.031 (.012)**	-0.023 (0.081)**	-0.091 (0.032)**	-0.073 (0.022)**	-0.073 (0.024)**	-0.073 (0.026)**	-0.11 (0.035)*	-0.057 (0.021)**
Sex	0.24 (0.15)+	0.12 (0.11)	0.53 (0.37)	0.50 (0.25)*	0.50 (.27)*	0.50 (0.28)+	0.59 (0.41)+	0.34 (0.25)
Married	0.051 (0.17)	0.039 (0.12)	0.21 (0.43)	0.12 (0.27)	0.12 (.30)	0.12 (0.33)	0.20 (0.51)	0.13 (0.32)
# Children	-0.66 (0.041)	-0.040 (0.028)	-0.16 (0.11)	-0.13 (0.074)+	-0.13 (0.081)	-0.13 (0.082)	-0.16 (0.13)	-0.11 (0.074)
'91 Absences *Sex	-0.18 (0.16)	-0.0059 (0.16)	-0.30 (0.37)	-0.31 (0.19)+	-0.31 (.21)	-0.30 (0.29)	-0.33 (0.43)	-0.21 (0.27)
R^2	0.19	0.18	-	-	-	-	-	-
Pseudo R^2	-	-	0.22	0.23	0.23	0.23	0.23	0.23
Log Likelihood	-229.82	-243.85	-201.89	-179.44	-179.44	-179.44	-176.15	-175.98

Note: N = 195
\+ $p < .10$
* $p < .05$
** $p < .01$

1. Sigma equals 1.69, with a standard error of 0.16, and is significant at p < .0l.
2. Overdispersion, e, equaled 1.20, and is not statistically significant.
3. Alpha equals 0.054, with a standard error of 0.20, and is not statistically significant.

statistics.

Perhaps some light is shed on this divergence by examining the log-likelihood values. The Tobit, Poisson, Negative Binomial, Ordered Logit, and Ordered Probit models are all computed by maximizing this value. Although OLS models do not use this method, it is possible to compute a log likelihood (e.g., Greene, 1992):

$$-\text{Log } L = -(N/2)\,[\,1 + 1n2\pi + 1n\ e'e\ /\ N)$$

N= Number of cases

e = Vector of residuals

The log likelihood can be computed for the transformed OLS regression, only the residuals need to be recalculated to reflect the difference between the square of the predicted term and the actual expected number of absences.

For modeling excused absences, as might be expected, the OLS regressions had the lowest (i.e., worst) log likelihoods. The Poisson regression also had a comparatively poor log-likelihood value. The remaining models had similar log-likelihood values. When modeling unexcused absences, where overdispersion was not statistically significant, the Poisson, Negative Binomial, and ordinal models had comparable log likelihood values, OLS regressions were still worst, and the Tobit model was somewhat worse than average.

Based on the regressions predicting the number of Type I errors described earlier, it is possible to estimate the propensity of these models for Type I errors given these samples' characteristics. Although further simulation is needed to make these estimates more accurate, the results are informative. Based on the skew, kurtosis, and sample size of excused absences (reported in Table 4), and assuming the independent variables are normally distributed, Probit is the most conservative model in this case. OLS, transformed OLS, Logit, and Negative Binomial, are also more conservative than the alpha levels suggest. Tobit, Overdispersed Poisson, and Poisson models are expected to yield more false positives than expected due to chance. For unexcused absences, the Probit, Negative Binomial, and OLS regression models are conservative. The Tobit and Poisson models are expected to produce too many Type I errors. The Logit, transformed OLS, and Overdispersed Poisson models are expected to produce Type I errors as frequently as suggested by the alpha levels.

With an understanding of the probability of Type I errors, and given the various log-likelihood values from the models, it is possible to make some conclusions regarding the significance of the results reported in Tables 5 and 6. For both excused and unexcused absences, prior absences is a highly significant predictor of future absences. It may also be concluded that tenure is significantly negatively related to absences in both cases, even though some models do not show a significant relationship between tenure and excused absences. The significance of sex is questionable. It is likely that the high significance shown by the Poisson model is a Type I error; however, as the simulation shows the Negative Binomial model to be conservative, the marginal effect discovered here should be given some weight. The interaction term also receives some mixed signals of significance, although these results seem more questionable.

Conclusions

The implications from these analyses are not entirely clear. What is clear, though, is that the type of analysis impacts observed results, and characteristics of what is being studied affects the accuracy of the type of analysis. The results of this study, although not providing a decisive best analysis strategy, do provide some insights that may benefit future absenteeism research, and perhaps more generally, benefit research involving count data.

The simulation results suggest that OLS regression is not overly sensitive to false positives. While the use of OLS regression with absenteeism data may still yield false negatives, we can conclude from this study that researchers of absenteeism do not need to ignore the significant findings from previous analyses.

Another implication of this study is that researchers should not use Poisson regression to analyze absenteeism data. Indeed, these results indicate that the Poisson regression can lead to a large number of false positives. At the very least, researchers would be better off even using OLS regression, but Overdispersed Poisson regression (see Gardner et al., 1995) seems promising as both an easy and accurate way of modeling count data. Despite previous results suggesting that the Tobit model is more sensitive than OLS (Baba, 1990; Hammer & Landau, 1981), these results suggest this sensitivity may be Type I errors. Finally, Negative Binomial seems a promising method of analyzing absence data. The model is very conservative in that it frequently yields false positives significantly fewer times than expected based on the alpha level. Although this may mean Negative Binomial models are prone to Type II errors, researchers may have more faith in findings that this model says are statistically significant.

The use of multiple methods also receives some support. Although the availability of methods depends on the statistics package being employed, it was relatively simple to perform all the analyses shown here. Comparisons across models, both in terms of their coefficients and log-likelihoods, may provide substantial information in cases where the choice of appropriate methodology is unclear.

Hopefully, future absenteeism research will move beyond simple OLS regression. The Overdispersed Poisson regression model is a promising alternative in that it is designed for count data, yields roughly the expected number of false positives, and is easily estimated. The

Negative Binomial models may be appropriate as a conservative test of significance. As suggested by other researchers (Baba, 1990; Hammer & Landau, 1981), the use of multiple methods of analysis appears to hold some hope for coping with the difficulties of absenteeism research. Used alone, or as part of a series of models, the Overdispersed Poisson regression and Negative Binomial model may help clarify hypothesis tests.

Future research should address the methodological issues of absenteeism research by exploring other potential models or methods of model comparisons, which perhaps may lead to a definitive best way to analyze absenteeism data. More extensive simulation may provide researchers with a better understanding of which models are most appropriate for certain circumstances, which could have implications beyond the domain of absenteeism. Additionally, more thought needs to be given to the issue of the sensitivity of these models to Type II errors. But perhaps most importantly, research on absenteeism should begin using additional, or perhaps alternative methodologies to perform hypothesis tests. The results from this paper confirm that previous empirical findings are still valid; however, this piece also shows that a number of alternatives to OLS regression exist which may help further the understanding of the correlates or causes of absenteeism.

References

Arbous, A. G., & Sichel, H. S. (1954). New techniques for the analysis of absenteeism data. Biometrica, 41, 77-90.

Amemiya, T. (1981). Qualitative Response Models: A Survey. Journal of Economic Literature, 19, 481-536,

Baba, V. V. (1990). Methodological issues in modeling absence: A comparison of least squares and Tobit analysis. Journal of Applied Psychology, 75, 428-432.

Blau, G., & Boal, K. (1989). Using job involvement and organizational commitment interactively to predict turnover. Journal of Management, 15, 115-127.

Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data: Comparisons and applications of some estimators and tests. Journal of Applied Economics, 1, 29-53.

Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the poisson model. Journal of Econometrics, 46, 347-364.

Clegg, C. W. (1983). Psychology of employee lateness, absence, and turnover: A methodological critique and an empirical study. Journal of Applied Psychology, 68 88-101.

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

Gibson, J. (1966). Toward a conceptualization of absence behavior of personnel in organizations. Administrative Science Quarterly, 11, 107-133.

Greene, W. H. (1992). LIMDEP, Castle Hill, Australia: Econometric Software.

Greene, W. H. (1993). Econometric Analysis, New York: Macmillan Publishing.

Gurmu, S. (1991). Tests for detecting overdispersion in the positive regression model. Journal of Business and Economic Statistics, 9, 215-222.

Gurmu, S., & Trivedi, P. K. (1992). Overdispersion tests for truncated Poisson regression models. Journal of Econometrics, 54, 347-370.

Hammer, T. H., & Landau, J. (1981). Methodological issues in the use of absence data. Journal of Applied Psychology, 66, 574-581.

Harrison, D. A., & Hulin, C. L. (1989). Investigations on absenteeism: Using event history models to study the absence-taking process. Journal of Applied Psychology, 74,300-316.

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. Econometrica, 52, 909-938.

Hulin, C. L. (1991). Adaptation, persistence, and commitment in organizations. In M. D. Dunnette & L. M. Hough (Eds.) Handbook of Industrial and Organizational Psychology (Vol. 2, pp. 445-505). Palo Alto, CA: Consulting Psychologists Press.

Hulin, C. L., & Rousseau, D. M. (1980). Analyzing infrequent events: Once you find them your troubles begin. In K. H. Roberts & L. Burstein (Eds.), Issues in Aggregation: New Directions for Methodology of Social and Behavioral Science (Vol. 6). San Francisco: Jossey-Bass.

Ivancevich, J. M. (1985). Predicting absenteeism from prior absence and work attitudes. Academy of Management Journal, 28, 219-228.

Johnson, R. A., & Wichern, D. W. (1992). Applied multivariate statistical analysis, Englewood Cliffs, NJ: Prentice Hall.

Kuhn, T. S. (1970). The Structure of Scientific Revolutions, 2nd Ed., Chicago, IL: University of Chicago Press.

Landy, F. J., Vasey, J. J., & Smith, F. D. (1984). Methodological problems and strategies in predicting absence. In P. S. Goodman & R. S. Atkin (Eds.), Absenteeism: New Approaches to Understanding, Measuring, and Managing Employee Absence, San Francisco, CA: Jossey-Bass.

Lee, L. F. (1986). Specification test for Poisson regression models. International Economic Review, 27, 689-706.

Leigh, J. P. (1985). The effects of unemployment and the business cycle on absenteeism. Journal of Economics and Business, 37, 159-170.

Maddala, G. S. (1984). Limited-dependent and qualitative variables in econometrics. Cambridge, England: Cambridge University Press.

Martocchio, J. J., & Harrison, D. A. (1993). To be there or not to be there?: Questions, theories, and methods in absenteeism research. Research in Personnel and Human Resources Management, 11, 259-328.

Martocchio, J. J., & Judge, T. A. (1994). A policy-capturing approach to individuals' decisions to be absent. Organizational Behavior and Human Decision Processes, 57, 358-386.

Mikalachki, A. & Gandz, J. (1982). Managing Absenteeism, Ontario, Canada: University of Western Ontario.

Morgan, L., & Herman, J. (1976). Perceived consequences of absenteeism. Journal of Applied Psychology, 61, 738-742.

Nicholson, N. (1977). Absence behavior and attendance motivation: A conceptual synthesis. Journal of Management Studies, 14, 231-252.

Nunnally, J. M. (1978). Psychometric Theory, New York: McGraw-Hill.

Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. Journal of Applied Psychology, 78, 569-582.

Rhodes, S. R., & Steers, R. M. (1990). Managing Employee Absenteeism, Reading, MA: Addison-Wesley.

Rushmore, C. H., & Youngblood, S. A. (1979). Medically-related absenteeism: Random or motivated behavior? Journal of Occupational Medicine, 21, 245-250.

Steers, R. M., & Rhodes, S. R. (1978). Major influences on employee attendance: A process model. Journal of Applied Psychology, 62, 319-327.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica, 26, 24-36.

Waters, L. K., & Roach, D. (1979). Job satisfaction, behavioral intention, and absenteeism as predictors of turnover. Personnel Psychology, 32, 393-397.

Watson, C. J., Driver, R. W., & Watson, K. D. (1985). Methodological issues in absenteeism research: Multiple absence measures and multivariate normality. Academy of Management, 10, 577-586.

Zavoina, R., & McElvey, W. (1975). A statistical model for the analysis of ordinal level dependent variables. Journal of Mathematical Sociology, Summer, 103-12