Pergamon

Journal of Safety Research, Vol. 28, No. 3, pp. 177–186, 1997

0022-4375/97 $17.00 + .00

PII S0022-4375(97)-00007-8

Modeling Fatal Injury Rates Using Poisson Regression: A Case Study of Workers in Agriculture, Forestry, and Fishing

A. J. Bailer, L. D. Reed, and L. T. Stayner

Injury surveillance data serves as the foundation of many safety studies. These studies frequently gather information on the number of injuries along with the number of employees at risk of injury in each of several strata where the strata are defined in terms of a series of important predictor variables. It is common for analyses of such data to examine injury rates separately for each predictor variable. The analysis of the *crude* or *unadjusted* injury rates give an overall indication of injury rate changes as a function of a particular predictor variable; however, further insights may be gained from analyses using Poisson regression models.

Poisson regression models are described as a means of analyzing rates adjusting for one or more predictor variables. In these models, the log rate of injury is expressed as a linear function of predictor variables. The interpretation of model parameters is given along with a presentation of the basic formulation of such models. Testing for trend, evaluation of confounding, and effect modification are illustrated using surveillance data describing occupational fatal injury rates as a function of year (1983–1992), gender and age for White workers employed in agriculture, forestry, or fishing. Data for this analysis were obtained from two sources: the National Traumatic Occupational Fatality (NTOF) database from the National Institute for Occupational Safety and Health provided counts of the fatal injuries, while data from the U.S. Bureau of Labor Statistics (BLS) provided counts on employment. Using an unadjusted trend model, a statistically nonsignificant decline in fatal injury rates over 1983–1992 is observed. Further analysis using Poisson regression revealed an interaction between gender and calendar year with males experiencing a weak, albeit significant, decrease and females experiencing a strong and significant increase. © 1997 National Safety Council and Elsevier Science Ltd

John Bailer, Ph.D., earned a doctorate in Biostatistics with an area of concentration in environmental health. Since 1988, he has been a professor in the Department of Mathematics and Statistics at Miami University in Oxford, Ohio. Since 1990, he has worked in conjunction with risk assessment researchers at the National Institute for Occupational Safety and Health. Dr. Bailer has published widely in the areas of occupational and environmental risk estimation with special interests in the design and analysis of occupational and environmental research studies.

Leslie Stayner, Ph.D., received his doctorate in Epidemiology from the University of North Carolina in 1989, and a Masters Degree in Epidemiology from the Harvard School of Public Health in 1980. He has been working with the National Institute for Occupational Safety and Health for approximately 15 years, originally involved in the conduct of industry-wide research projects, and is currently involved in managing their Risk Assessment Program.

Larry Reed, M.S., earned a Master's degrees in Preventive Medicine and in Industrial and Systems Engineering. He has been employed at the National Institute for Occupational Safety and Health since 1977. In his current position, Mr. Reed manages the development of NIOSH policy statements to regulatory agencies and document development including criteria documents, alerts and hazard controls. Prior to these responsibilities, he worked in the research areas of personal protective equipment, industrial hygiene, and engineering control technology. He also teaches mathematics on a part-time basis in the College of Evening and Continuing Education at the University of Cincinnati.

INTRODUCTION

The basic data for many occupational safety studies are injury rates. The number of injuries and the corresponding number of workers employed are recorded for several levels of one or more predictor variables, such as industry or occupation. As an example, consider an analysis of occupational fatal injury data. One might be interested in assessing research hypotheses concerning whether the injury rates differ as a function of workplace characteristics such as industry, job title, shift, and worker characteristics such as gender, race, and age. Prior to any analysis, the number of workers employed (or time at risk) and the number of fatal injuries would be recorded for each unique workplace-worker combination (i.e., for each strata defined by industry-job title-shift-gender-race-age combinations). Once these data are gathered, it is common for the analyses to examine injury rates separately for the levels of each predictor variable. This type of analysis is sometimes called a *crude* or *unadjusted* analysis of rates since the potential confounding effects of other predictor variables are ignored. These unadjusted analyses provide valuable information regarding the overall public health impact of certain variables on injury rates; however, these analyses may mask important clues related to occupational fatal injury etiology.

Further insight into the relationship between injury rates and the predictor variables may be obtained by using regression methods. These methods may be used to assess whether or not a trend in rates is statistically significant. In addition, these methods may be used to assess whether the association between a risk factor of interest and the rate of adverse response is independent of other factors, that is whether there is confounding. While an analysis of crude rates within strata defined by potentially confounding factors partially addresses this question, the number of strata increases dramatically as the number of potential confounders that are simultaneously considered grows large. Thus, techniques for the assessment and simultaneous control of a large number of confounding variables are desirable. It is also possible that the strength and/or direction of the relationship between the risk factor of interest and the rate of adverse response may differ in strata defined by levels of other predictor variables, so-called effect modification. For example, as we will illustrate shortly, the trend in fatal injury rates is much different for males and females. These three issues — testing for trend, evaluating confounding, and detecting effect modification — provide the motivation for a consideration of regression methods as an alternative to univariate methods for the analysis of injury data in this paper. To illustrate these issues, we use Poisson regression modeling methods to address questions of a calendar year trend in fatal injury rates for workers employed in the agriculture, forestry, and fishing industries.

MATERIAL AND METHODS

Data

Workers in agriculture, forestry, and fishing are employed in an industry with nearly the highest occupational fatal injury rates with annual fatal injury rates of approximately 20 per 100,000 workers documented over the years 1980–1989 (NIOSH, 1993). These rates are seen to vary with age within agriculture, forestry, and fishing (see, for example, NIOSH, 1993). For the analyses that follow, we restricted our attention to workers that were identified as being in the agriculture, forestry, and fishing industry division (SIC codes 01-09 from Office of Management and Budget, 1987) and "farmers/foresters/fishers" occupation group (U.S. Department of Commerce, 1982). We further restricted our analysis to White workers, the largest racial category in our data set, to simplify our analysis. We selected this combination of industry/occupation/race categories for exposition — other industry/occupation/race categories could have been selected. (As an aside, the total number of fatal injuries in non-Whites averaged approximately 52 deaths each year during this interval, while the annual average number employed in this category was approximately 211,110. Unadjusted non-White fatal injury rates ranged from 18 to 30 deaths per 100,000 workers with no statistically significant trend observed.)

Counts of deaths in this cohort were cross-classified by calendar year (1983–1992), gender (male, female), and age category (11 categories — 16–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, and 65+). These counts were obtained from the National Institute for Occupational Safety and Health National Traumatic Occupational Fatality (NTOF) database, a death certificate-based registry, and were combined with data on employment from the U.S. Bureau of Labor Statistics (BLS). The BLS employment data were based upon unpublished data tabulated from the current population survey (U.S. Bureau of the Census, 1978) constructed in response to an interagency agreement between BLS and NIOSH.

Data Description

We selected this combination of industry and occupation categories for exposition of the models. Agriculture, forestry, and fishing represent a group with large numbers of individuals employed along with fairly high crude fatal injury rates. Agriculture, forestry, and fishing represent a group with large numbers of individuals employed — from 2,850,803 in 1983, and decreasing to 2,581,603 in 1992. Approximately 82.5% of the workforce is male, with this proportion remaining fairly stable over the last decade. Small changes are being experienced in this workforce age distribution over the years 1983 through 1992. In particular, the proportion of the workforce in the lowest three age categories had each declined by 2.4 to 3.0%, while the percentage of workers in the highest age category has increased by approximately 1.0%.

Statistical Analyses

Objectives in the analysis of rates include a determination of which predictor variables are related to the rates and how these predictor variables are functionally or mathematically related to the rates (Breslow & Day, 1987). In short, we model rates as some explicit function of relevant predictor variables. One framework in which the analysis of rates may be conducted is called a Poisson regression model. This model has frequently been applied to the analysis of cohort data in epidemiology (Berry, 1983; Frome, 1983; Frome & Checkoway, 1985; Whittemore, 1985), and is now described in many books (Breslow & Day, 1987; Checkoway, Pearce, & Crawford-Brown, 1989; Selvin 1995). In our presentation below, we focus on a quick overview of these methods with special attention toward their use in the analysis of fatal injury statistics. Interested readers are encouraged to examine the references given above for greater detail on the use of these techniques in general. The methods we describe in this section serve as alternatives to the stratified analyses that are often encountered in injury epidemiology. In those analyses, injury rates are explored separately for strata defined by combinations of predictor variables. For example, 220 separate strata are defined by the 10 years × 2 levels of the gender variable × 11 age categories. Rates defined specifically for each strata may be unstable due to small numbers of observations in each stratum (Checkoway et al., 1989, Section 3.2). Thus, when many predictor variables are to be simultaneously controlled, alternatives to such analyses should be considered. Statistical models provide a natural means of addressing such a problem.

The injury rates are estimated by the number of injuries (d) divided by the number of person-years at risk (N), i.e.,

$$rate = (number\ dead)/(person\text{-}years\ at\ risk).$$

The number dead, d, is assumed to follow a Poisson distribution with mean $\mu_d = N * \lambda$ where λ is the true underlying injury incidence rate. The mean number of deaths, or ultimately, the rate, is then modelled as some parametric function of the predictor variables. Since a Poisson variate is assumed to have a mean > 0, this function of the predictors is frequently constrained so that the values of this function are greater than zero. A common choice for modeling the rate λ, as a function of a set of predictor variables $X_1, \ldots, X_p$, is a log-linear model of the following form:

$$\log(\lambda) = \log(\mu_d/N) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p \quad (1a)$$

or, equivalently,

$$\log(\mu_d) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \log(N) \quad (1b)$$

where β_i is the regression coefficient relating the mean number of injuries to the i^{th} predictor variable X_i ($i = 1, \ldots, n$). The log transformation ensures that the model-based predictions of rates are constrained to be greater than or equal to zero.

Thus, log(rate) is a linear function of the predictor variables. In the language of generalized linear models (GLIMs), log is a "link" function and log(N) is called the "offset" (Dobson, 1990; McCullagh & Nelder, 1989). As we see in equation 1b, the log function "links" the average number of deaths to the linear combination of the predictor variables while the log(N) term enters the model with fixed weighting coefficient (a beta of 1 in this case). The analyses presented herein are conditional on the employment counts. Other denominators, N, might be of interest. For example, the number of hours employed could be used as an alternative to the employment counts. It is not hard to imagine that a 16-year-old working part-time does not experience the same amount of time at risk as a 35-year-old working full-time. In addition, holders of multiple jobs might be accomodated in this analysis by only counting the hours worked in a particular industry of interest. The rate analysis described herein can easily accomodate this change, where now the injury rates are expressed as fatal injuries per person-hours of work.

The choice of a Poisson distribution for the number of deaths, d, leads to the variance function in the GLIM in which the variance of the response is equal to the mean response. Notice that this represents a clear point of departure from the multiple linear regression model in which equal (homogenous) variances are assumed. Count data, by its very nature, immediately violates this multiple regression assumption, leading to the need to consider more complicated regression models such as GLIMs. This model structure can be fit using appropriate software (e.g., SAS, 1993) and hypothesis tests and parameter estimates, $\hat{\beta}$s, along with associated standard errors can be generated. As a technical aside, these estimates and standard errors are obtained by means of maximum likelihood estimation often implemented in GLIMs in an iteratively reweighted least squares algorithm (see McCullagh & Nelder, 1989 for more detail).

Critical assumptions are made in this modeling exercise. First, a correct specification of the link function is important (i.e., are the log rates linearly related to the predictor variables of interest?). The adequacy of the model specification can be assessed by residual analyses (see, for example, Sec. 2.4, McCullagh & Nelder, 1989). Second, the Poisson response distribution implies that the variance is equal to the mean, which is an important assumption, and if violated, the standard errors associated with the parameter estimates are incorrect. The most common violation of this assumption is for the variance to exceed the mean — so-called "overdispersion." Methods for incorporating overdispersion can be found in a variety of sources, including McCullagh and Nelder (1989, section 6.4) and SAS (1993). As an aside, adjustments for overdispersion have been incorporated into previous analyses of trends in injury rates, in particular, childhood drowning rates (Brenner, Smith, & Overpeck, 1994). A global check of the model specification, both mean and variance components, is contained in the χ^2 statistic, which is constructed from the observed and expected counts (Dobson, 1990, Section 9.5). Finally, a comparison of statistical models based upon a measurement of the discrepancy of a model's fit, the "deviance," is also used to qualitatively evaluate the adequacy of a Poisson regression model. In particular, if the deviance is approximately equal to the model degrees of freedom, then the model is considered adequate.

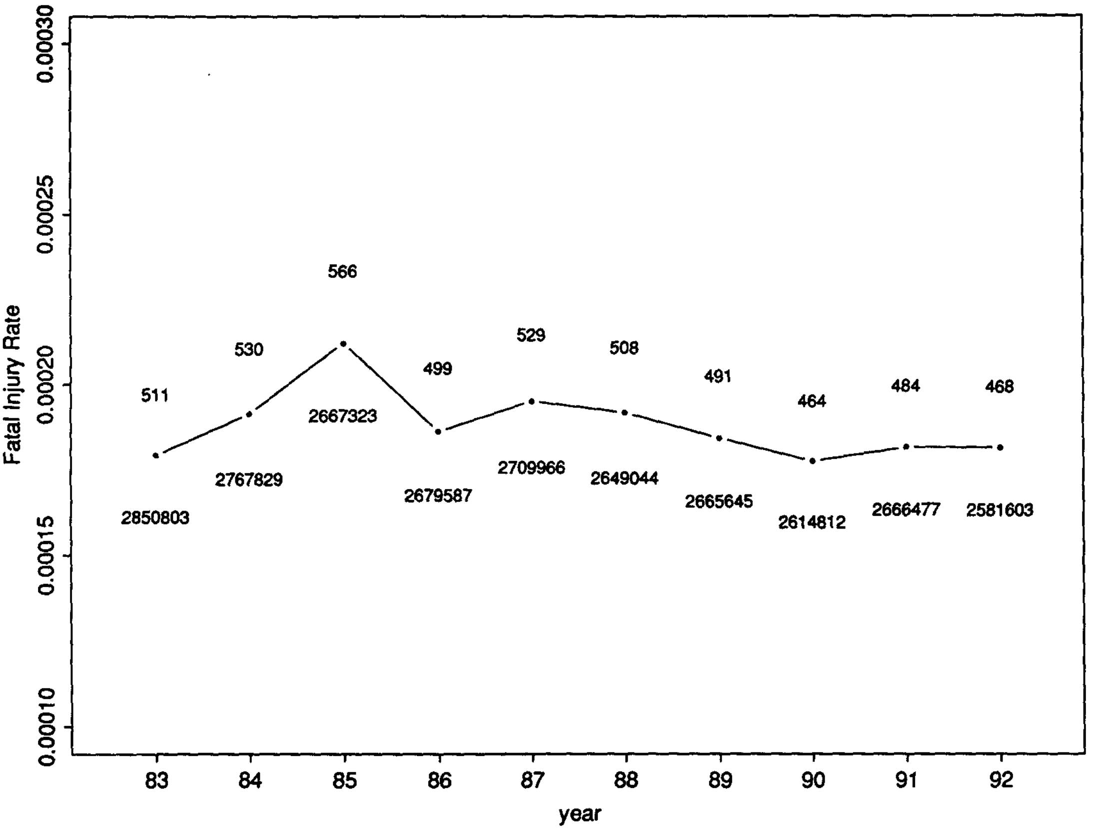
The predictor variables in the model above might be categorical, such as gender. As an example, if X_1 is defined as $X_1 = 1$ if female and $X_1 = 0$ if male (i.e., X_1 is an indicator variable), then the

$$(\textit{injury rate for females}) = exp(\beta_1) \times (\textit{injury rate for males})$$

where the injury rate for males is $\exp(\beta_0)$ and all other predictor variables are held constant. This leads to another common interpretation of the parameters of these models. The injury incidence *rate ratio* (RR) for females as compared to males is $\exp(\beta_1)$. An estimate of this rate ratio is given by $\exp(\hat{\beta}_1)$ where $\hat{\beta}_1$ is the point estimate of β_1. An approximate $100 \times (1 - \alpha)\%$ confidence interval for this rate ratio is given by $\exp[\hat{\beta}_1 \pm Z_{\alpha 2}\ \text{s.e.}(\hat{\beta}_1)]$ where $Z_{\alpha/2}$ is the upper-tailed critical value from a standard normal distribution (e.g., $Z_{.025} = 1.96$) and s.e.$(\hat{\beta}_1)$ is the standard error associated with $\hat{\beta}_1$.

All model fits reported in the results section were obtained using the SAS GENMOD procedure (SAS, 1993). As a final caveat, many important issues need to be considered when utilizing these models. For example, are these models correctly specified — both in terms of the predictor variables considered and the response distribution. Although not presented in the text, we conducted additional analysis to assess the goodness-of-fit of the regression models we employed. In addition, the quality of the data, such as undercounting of certain ages or of a particular gender should be evaluated. Finally, the death certificate (numerator) data is based on "usual" industry or occupation while the survey (denominator) employment data, reflects the actual job in the week of the

FIGURE 1
PLOT OF THE OCCUPATIONAL FATAL INJURY RATES FOR AGRICULTURE/FORESTRY/FISHING INDUSTRY AND OCCUPATION DIVISIONS TABULATED FOR YEARS 1983 THROUGH 1992. EACH POINT IS ANNOTATED WITH THE NUMBER OF FATALITIES (ABOVE THE OBSERVED INJURY RATE) AND THE NUMBER EMPLOYED (BELOW THE OBSERVED INJURY RATE)



survey. This adds yet another potential confounding factor to the analyses of rates based upon different data sources.

RESULTS

Unadjusted Trends in Fatal Injury

The total number of fatal injuries and number employed was classified by year, and the injury rates by this classification are plotted in Figure 1. From this figure, it appears that the occupational fatal injury rate has remained relatively constant over the time period 1983 through 1992. In this figure a rate of 0.00020 corresponds to 2 fatal injuries per 10,000 workers or 20 fatal injuries per 100,000 workers.

A simple univariate Poisson regression model may be use to formally test whether or not there is evidence for a trend in the rates as a function of one predictor variable, namely calendar year. For this first model, the raw data would be the number dying in occupational fatal injuries and the number employed for each year. Thus, no adjustment for other predictor variables are considered in this model. Formulaically, this is represented as

$$log(\lambda_{year})=\beta_0+\beta_1(\text{YEAR}) \qquad (2)$$

where λ_{year} is the injury rate for each year. This model assumes that the injury rate is the same for all white agriculture/forestry/fishing workers regardless of gender and age. The raw data for fitting such a model are presented in Figure 1. The number of fatalities in each year is displayed above the observed injury rate while the size of the workforce is displayed below this rate.

The regression model described in equation 2, hereafter referred to as "model (2)," implies that a plot of log(rate) versus year should be linear with slope β_1. As an aside, year was rescaled, YEAR = calendar year − 1983, for this model and all later models so that the intercept parameter represents the log rate for 1983.

The coefficients from fitting model (2) were $\hat{\beta}_0 = -8.5469$ and $\hat{\beta}_1 = -0.0072$ with s.e.$(\hat{\beta}_0) = 0.0256$ and s.e.$(\hat{\beta}_1) = 0.0049$. This model suggests that the log-transformed injury rate decreases −0.0072 units per year. Alternatively, the estimated change in fatal injury rates from 1 year to the next was exp(−0.0072) = 0.993, which translates into an annual decrement in the fatal injury rate of about 0.7% [= 100 × (1−0.993)%]. An approximate 95% CI for the parameter associated with year is −0.0072 ± 1.96(0.0049) = (−0.0168, 0.0024). Since $\beta_1 = 0$, is contained in this interval, no significant change in crude fatal injury rates per year is observed. This observation is corroborated by the p-value associated with a test of $\beta_1 = 0$, which was 0.14.

Assessing Confounding

Other natural questions in the analysis of trends in fatal injury rates are whether the observed trend in crude fatal injury rates with year is the same when controlled for other predictor variables. If the relationship between injury rates and year adjusted for other predictor variable differ from the relationship between crude injury rates and year, then evidence of confounding is present (Kleinbaum, Kupper, & Morgenstern, 1982). The Poisson regression model can be used to evaluate these questions. For this discussion, (calendar) year is viewed as the risk factor of interest while gender and age will be considered potential confounders. The following model may be used to assess whether the relationship between injury rates and calendar year is confounded by age and gender:

$$log(\lambda[year\text{-}gender\text{-}age]) = \beta_0 + \beta_1(\text{YEAR}) + \beta_2 \text{I}(\text{GENDER}=female) + \sum_{i=1}^{10} \alpha_i \text{I}(\text{AGE}=i) \quad (3)$$

where $\lambda_{[\text{year-gender-age}]}$ is the injury rate for a particular year-gender-age combination, I(GENDER = female) = 1 if female and = 0 if male and I(AGE = 1) = 1 if the first age category and = 0 otherwise. Other age categories are similarly defined with the last age category serving as a reference category. This model suggests that the relationship between injury rate and year is the same for both genders and all age categories. This model allows for differences in the baseline injury rates for the three variables YEAR, GENDER, and AGE. For example, the log-injury rate for 16–19-year-old females in 1983 would be $\log(\lambda_{[83\text{-female-age} = 1]}) = \beta_0 + \beta_2 + \alpha_1$ implying the injury rate is $\lambda_{[83-\text{female}-\text{age}=1]} = \exp(\beta_0 + \beta_2 + \alpha_1)$. The interpretation of β_1 is now the change is log-transformed injury rate associated with a 1 year change when gender and age are held constant. The raw data (not shown) for fitting such a model is the number of fatalities and the number of workers in each year-gender-age combination. If the estimate of β_1, the regression coefficient associated with the effect of year, changes appreciably from fitting model (2) versus fitting model (3), this suggests a confounding of the injury rate — year relationship by the gender and age predictor variables.

To assess confounding, we fit model (3) in which age (entering the model as 10 categorical variables) and gender (entering the model as 1 categorical variable) are added to a model containing calendar year. The results of this model fit is given in Table 1. Note, in particular, that the parameter estimate associated with calendar year has changed from −0.0072 from the crude analysis to −0.0095 (with standard error of 0.0049). We see some evidence of confounding since the point estimate of the year coefficient changes by 32% (−0.0095/−0.0072). Finally, while there is still no strong evidence of significant trend in fatal injury rates with calendar year, the p-value associated with the test of the parameter associated with year in a model with gender and age considered is 0.05 in contrast to the p-value of 0.14 observed in the analysis of unadjusted rates. This suggests that the simultaneous control of other predictor variables appears to have strengthened the evidence for decline in the fatal injury rates with calendar year.

Assessing Effect Modification

Further evaluation of the data might involve questions about whether the trend with calendar

TABLE 1
PARAMETER ESTIMATES, STANDARD ERRORS AND P-VALUES RESULTING FROM THE FIT OF MODEL(3)—ASSESSING CONFOUNDING

Model Term	Level of Categorical Variables	Parameter Estimates	Standard Errors	P-values
INTERCEPT		-7.4297	0.0353	0.0000
AGE	16-19	-2.0910	0.0873	0.0000
	20-24	-1.3007	0.0561	0.0000
	25-29	-1.1991	0.0539	0.0000
	30-34	-1.2532	0.0567	0.0000
	35-39	-1.2036	0.0592	0.0000
	40-44	-1.0366	0.0608	0.0000
	45-49	-0.9880	0.0614	0.0000
	50-54	-0.8195	0.0587	0.0000
	55-59	-0.7473	0.0552	0.0000
	60-64	-0.6755	0.0546	0.0000
	65 and up*	0.0000	0.0000	
GENDER	Female	-2.6930	0.1213	0.0000
	Male*	0.0000	0.0000	
YEAR		-0.0095	0.0049	0.0516

* These levels are the reference levels for the categorical variables.
Note: P-value = 0.0000 implies that the calculated P-value <0.0001.

year varies by different levels of the other covariates. If the trend differs within levels of other predictor variables, then evidence of *effect modification* is present (Kleinbaum et al., 1982).

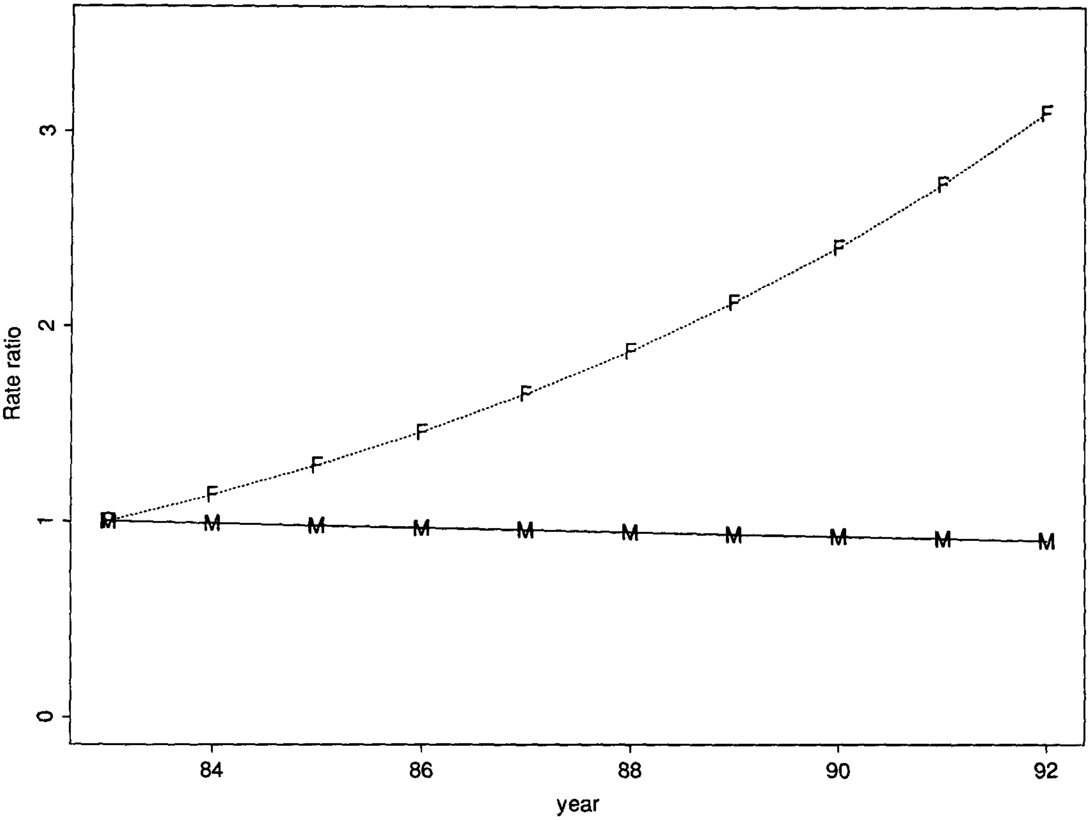
The following model may be used to assess whether the predictor variables for calendar year varies by gender or age:

$$log(\lambda_{[year\text{-}gender\text{-}age]}) = \beta_0 + \beta_1(\text{YEAR}) + \beta_2\text{I(GENDER}=female) + \beta_3\text{YEAR*I(GENDER}=female) + \sum_{i=1}^{10}\alpha_i\text{I(AGE}=i) + \sum_{i=1}^{10}\gamma_i\text{YEAR*I(AGE}=i) \quad (4)$$

where β_3 represents how the trend for females differs from that for males and γ_i represents how the trend for age category i differs from that for the reference age category. Even though the raw data for this model is identical to that for model (3), the interpretation of the parameters differs dramatically. As noted previously, Model (3) implicitly assumes that the effect of year on log-injury rates is the same for males and females and for all ages (i.e., a no interaction assumption — $\beta_3 = \gamma_1 = \ldots = \gamma_{10} = 0$). Model (4) explicitly incorporates interaction or effect modification terms in the model via the cross-product terms YEAR*I(GENDER = female) and YEAR*I(AGE = i). If the β_3 or the γ_i parameters differ significantly from zero, then evidence is present that gender or age is an effect modifier of the year effect on fatal injury rates. As an example, if $\beta_3 \neq 0$, then the log-transformed injury rates are predicted to change by β_1 with each passing year in males and by $\beta_1 + \beta_3$ with each passing year in females (holding age constant).

The issue of effect modification of the relationship between fatal injury rates and calendar year by age and gender was evaluated by fitting Model (4). This model was fit (not reported), and there was little evidence that age was an effect modifier of the calendar year effect, since the year-age effect did not differ significantly ($p = 0.57$) from zero. (Technical point: this p-value was obtained from multiple degree of freedom test of $\gamma_1 = \ldots = \gamma_{10} = 0$, which involved comparing discrepancy measures from nested models, differences in deviances. Interested readers should consult McCullagh and Nelder (1989) for details concerning the comparison of nested models.) In contrast, there was evidence that gender was an effect modifier of year since the "YEAR*GENDER" interaction appears to be significantly ($p = 0.02$) different from zero. This implied that the trend in females differs significantly from the trend in males. In additional analyses, we modified Model (4) by elim-

FIGURE 2
PLOT OF THE RATE RATIO FOR FATAL INJURY FOR AGRICULTURE/FORESTRY/FISHING INDUSTRY AND OCCUPATION DIVISIONS OVER YEARS 1983 THROUGH 1992. SEPARATE TRENDS ARE DISPLAYED FOR MALES (SYMBOL "M" WITH CONNECTING SOLID LINE) AND FOR FEMALES (SYMBOL "F" WITH CONNECTING DASHED LINE). NOTE THAT THE RATE RATIO IS IN REFERENCE TO THE MALE AND FEMALE RATES IN 1983.



inating the age-year interaction terms. This model is given by

$$log(\lambda_{[year\text{-}gender\text{-}age]}) = \beta_0 + \beta_1(\text{YEAR}) + \beta_2 \text{I}(\text{GENDER}=female) + \beta_3 year^* \text{I}(\text{GENDER}=female) + \sum_{i=1}^{10} \alpha_i \text{I}(\text{AGE}=i) \quad (5)$$

The fit of Model (5) is reported in Table 2 . This additional analysis also demonstrated significant evidence of effect modification by gender. In addition, greater precision in the estimation of the year and year-gender effects was achieved by dropping the year-age interactions.

The point estimate for the trend in males is −0.0113, while the point estimate for the trend in females is −0.0113 + 0.1369 = 0.1256. Both of these parameters differ significantly from zero. The test for a trend effect in males has an associated p-value of 0.0214, which can be obtained directly for the "YEAR" row of Table 2. This is because "male" is the reference level of the gender variable for this analysis. The test for a trend effect in females results in a p-value of .0040, which is easily obtained by re-running the Poisson regression analysis using "female" as the reference level for the gender variable (analysis not shown).

This relationship is illustrated in Figure 2 in which the rate ratios for males and females are displayed for years 1983 through 1992. This figure illustrates that while the males may have a slight decrease in fatal injury rates with cal-

TABLE 2
PARAMETER ESTIMATES, STANDARD ERRORS AND P-VALUES RESULTING FROM THE FIT OF A MODIFICATION OF MODEL (5)—ASSESSING EFFECT MODIFICATION OF GENDER ONLY

Model Term	Level of Categorical Variables	Parameter Estimates	Standard Errors	P-values
INTERCEPT		-7.4217	0.0353	0.0000
GENDER	Female	-3.3666	0.2682	0.0000
	Male	0.0000	0.0000	
AGE CATEGORY	16-19	-2.0912	0.0873	0.0000
	20-24	-1.3008	0.0561	0.0000
	25-29	-1.1995	0.0539	0.0000
	30-34	-1.2531	0.0567	0.0000
	35-39	-1.2034	0.0592	0.0000
	40-44	-1.0369	0.0608	0.0000
	45-49	-0.9880	0.0614	0.0000
	50-54	-0.8193	0.0587	0.0000
	55-59	-0.7476	0.0552	0.0000
	60-64	-0.6759	0.0546	0.0000
	65 and up	0.0000	0.0000	
YEAR		-0.0113	0.0049	0.0214
YEAR* GENDER	Female	0.1369	0.0439	0.0018
	Male*	0.0000	0.0000	

* These levels are the reference levels for the categorical variables.
Note: P-value = 0.0000 implies that the calculated P-value < 0.0001.

endar year, the females may be increasing. It is important to note that this figure is a rate ratio relative to the fatal injury rates in males and females in 1983. This rate for females is much lower as compared to males. In particular, the estimated 1983 rate for females is $\exp(-3.3666) = 0.0345$ the 1983 rate for males.

DISCUSSION

In this paper, we described a regression technique appropriate for the analysis of injury rate data. An analysis of fatal injury rates among workers in agriculture, forestry, and fishing industries and occupations is used to illustrate the use and interpretation of this technique. In this example, the analysis of crude or unadjusted injury rates showed a nonsignificant decrease in these rates over the time period 1983 to 1992. However, more detailed analyses involving other predictor variables suggested that the trend was confounded by the gender and age of the workers. In addition, an interaction between calendar year and gender suggested that men experienced a weak, but significant, decrease in contrast to the trend in women, which was a strong and significant increase. This differing pattern of trends would not have been noted using a crude or unadjusted analysis. As noted above, the crude analysis suggested no trend in fatal injury rates in this worker population. Thus, even though greater complexity in the statistical analysis of injury rate data is introduced by the use of Poisson regression models, a more detailed exploration of the simultaneous effects of many variables on these rates is possible.

ACKNOWLEDGEMENTS

The authors are grateful to Lynn Jenkins of NIOSH for her assistance in obtaining an electronic version of the mortality data and to Bob McIntire of the Bureau of Labor Statistics for his assistance in obtaining an electronic version of the employment data that served as the denominator in the rate modeling portion of this study. The critical review of this manuscript by Drs. Gordon Smith, Marcie Kresnow, Terry Wassell, Nancy Stout and William Halperin led to an improved presentation.

REFERENCES

Berry, G. (1983) The analysis of mortality by the subject-years method. *Biometrics 39*, 173–184.

Brenner, R. A., Smith, G. S., & Overpeck, M. D. (1994). Divergent trends in childhood drowning rates, 1971 through 1988. *Journal of the American Medical Association, 271*, 1606–1608.

Breslow, N., & Day, N. (1987). *Statistical methods in cancer research, Vol. II — The design and analysis of cohort studies.* IARC Scientific Pub. No. 82, IARC, Lyon, FR.

Checkoway, H., Pearce, N. E., & Crawford-Brown, D. J. (1989). *Research methods in occupational epidemiology.* New York: Oxford University Press.

Dobson, A. (1990). *An introduction to generalized linear models.* New York: Chapman & Hall.

Frome, E. L. (1983). The analysis of rates using poisson regression models. *Biometrics, 39*, 665–674.

Frome, E. L., & Checkoway, H. (1985). Use of Poisson regression models in estimating incidence rates and ratios. *American Journal of Epidemiology, 121*, 309–323.

Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic research principles and quantitative methods.* New York: Van Nostrand Reinhold.

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

NIOSH. (1993). *Fatal injuries to workers in the United States, 1980–1989: A decade of surveillance.* DHHS (NIOSH) Number 93-108S, Cincinnati.

Office of Management and Budget. (1987), *Standard industrial classification manual.* Washington, DC: Author.

SAS. (1993). *The GENMOD procedure.* SAS Tech. Report P-243. Cary, NC: SAS Institute.

Selvin, S. (1995). *Practical biostatistical methods.* Belmont, CA: Duxbury Press.

Whittemore, A. S. (1985). Analyzing cohort mortality data. *The American Statistician, 39*, 437–441.

U.S. Bureau of the Census. (1978). *The current population survey: design and methodology.* Technical Paper 40.

U.S. Department of Commerce, Bureau of the Census. (1982). *1980 census of population: Alphabetic index of industries and occupations.* Publication PHC80-R3.