Pergamon

Journal of Safety Research 36 (2005) 413 – 421

www.elsevier.com/locate/jsr

www.nsc.org

Analysis of non-fatal and fatal injury rates for mine operator and contractor employees and the influence of work location

Vijia K. Karra *

National Institute for Occupational Safety and Health,[1] Pittsburgh Research Laboratory, Cochrans Mill Road, P.O. Box 18070, Pittsburgh, PA 15236, USA

Received 17 January 2005; received in revised form 13 July 2005; accepted 30 August 2005
Available online 21 November 2005

Abstract

Introduction: Mining injury surveillance data are used as the basis for assessing the severity of injuries among operator and contractor employees in the underground and surface mining of various minerals. *Method:* Injury rates during 1983–2002 derived from Mine Safety and Health Administration (MSHA) database are analyzed using the negative binomial regression model. The logarithmic mean injury rate is expressed as a linear function of seven indicator variables representing Non-Coal Contractor, Metal Operator, Non Metal Operator, Stone Operator, Sand and Gravel Operator, Coal Contractor, and Work Location, and a continuous variable, RelYear, representing the relative year starting with 1983 as the base year. *Results:* Based on the model, the mean injury rate declined at a 1.69% annual rate, and the mean injury rate for work on the surface is 52.53% lower compared to the rate for work in the underground. With reference to the Coal Operator mean injury rate: the Non-Coal Contractor rate is 30.34% lower, the Metal Operator rate is 27.18% lower, the Non-Metal Operator rate is 37.51% lower, the Stone Operator rate is 23.44% lower, the Sand and Gravel Operator rate is 16.45% lower, and the Coal Contractor rate is 1.41% lower. Fatality rates during the same 20 year period are analyzed similarly using Poisson regression model. Based on this model, the mean fatality rate declined at a 3.17% annual rate, and the rate for work on the surface is 64.3% lower compared to the rate for work in the underground. With reference to the Coal Operator mean fatality rate: the Non-Coal Contractor rate is 234.81% higher, the Metal Operator rate is 5.79% lower, the Non-Metal Operator rate is 47.36% lower, the Stone Operator rate is 8.29% higher, the Sand and Gravel Operator rate is 60.32% higher, and the Coal Contractor rate is 129.54% higher.

Keywords: Mining; Injury; Fatal; Non-fatal; Regression

1. Introduction

Mining is one of the most hazardous industries in the world. Underground mining of coal presents the most dangerous work environment. Occupational risk in underground mining is much higher than in surface mining. The risk in mining of stone, sand and gravel, and other metallic and non-metallic minerals is lower than that for coal. Injuries, fatal or non-fatal, could result from dust and gases, fires, slips, falls, interaction with machinery, confined working spaces, repetitive work, vibrations, and so forth. The term *commodity* refers to application specific operator or contractor and the category *work location* refers to underground or surface operation. Injury rates are typically analyzed annually by surveillance personnel for each commodity and work location and data for 1983–2003 are available (National Institute for Occupational Safety and Health [NIOSH], 2004). No attempt, until now, is made to model the injury rate data for all commodities and work locations. Such an approach uses all the data in the entire period, providing an opportunity to include, if necessary, any potential interaction effects within the

* Tel.: +1 412 386 6648; fax: +1 412 386 6764.
E-mail address: vkarra@cdc.gov.

[1] The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

0022-4375/$ - see front matter
doi:10.1016/j.jsr.2005.08.002

covariates and their specific quantitative impact on the injury rate. The integrated model also sheds light on the relative impact of various commodities and the work location, which is a valuable input for development and field test planning of injury prevention interventions. It also provides an important measure of the annual injury rate reduction. Relative impact information is valuable to mine operators for targeting injury prevention in high risk commodities and work locations through appropriate resource allocation. Researchers can undertake further analysis of injury data within the identified high risk areas to understand accident/injury classification, source of injuries, and miner's activity at the time of accident, nature of injury, and so forth.

The nature of injury data is not continuous but discrete and count based. An appropriate model for discrete injury data covering several commodities and work locations is needed. In the area of modeling of injuries in mining and related industries, a literature search revealed the following. Indices of mine safety data such as the hazard rate, safety, risk, and mean time between accidents were assessed from the U.S. mine accident data for a 9.5 year period during 1975–1984. Both fatal and non-fatal injuries combined for one mine operation were analyzed using the Poisson distribution model (Kerkering & McWilliams, 1987). Injury severity in U.S. underground bituminous coal mines, during 1975–1981, was assessed relative to mine and miner characteristics using logistic regression (Bennet & Passmore, 1984). Here the focus was on underground bituminous coal only, and the logistic regression model is used for modeling a dichotomous dependent variable. Multiple regression was used for evaluating factors associated with occupational injury severity in New South Wales underground coal mining industry (Hull, Leigh, Driscoll, & Mandryk, 1996). For count based injury data, when the counts are large (>10), a normal distribution can be assumed and multiple regression can be used. With overdispersed and annually collected injury data, multiple regression models may not work due to problems such as autocorrelation. Maiti and Bhattacharya (1999) evaluated risk of occupational injuries among Indian underground coal workers through multinomial logit analysis. Risk indices for these workers were developed employing various personnel and workplace independent variables using logistic regression analysis (Maiti, 2003). Predictor variables included age, experience, occupation, location, and specific mine. Commingled fatal and non-fatal injury data were used.

Fatal injury rates during 1983–1992 among agriculture, forestry, and fishing industry workers were studied using categorical (indicator) and continuous (year) covariates and their interactions in Poisson regression models (Bailer, Reed, & Stayner, 1997). The methodology presented is very important to this study on miners' non-fatal injury analysis. Initial effort of the present study, therefore, is focused on evaluating the Poisson model.

In transportation research, Miaou (1994) studied the relationship between truck accidents and geometric design of road sections using both Poisson and negative binomial (NB) regressions and compared them. He cautioned using non-max-likelihood methods for fitting the NB model. His data had many zeros in the dependent variable responses and the selected model was a zero inflated Poisson (ZIP) model. The mining injury data did not have such zero responses. Effects of work zone presence on injury and non-injury crashes on California freeways were studied by Khattak, Khattak, and Council (2002) using the negative binomial model. They used commingled fatal/non-fatal injury and non-injury 1993 data from California. Their approach is adaptable for studying intervention effectiveness. Byers, Allore, Gill, and Peduzzi (2003) successfully employed the NB model for fitting overdispersed data in their aging research for evaluating intervention effectiveness among older persons. They recommend using the NB model over the Poisson model for overdispersed discrete outcomes.

In this paper, the focus is on non-fatal and fatal injury rates to assess their trend over the years and to be able to delineate the relative effects of commodity and work location variables using the Poisson or negative binomial statistical models.

2. Materials and methods

2.1. Data and data description

Non-fatal and fatal injury data were obtained from the Mine Safety and Health Administration (MSHA) database. In particular, data on *commodity*, *work location*, *non-fatal injuries*, and *employee hours* were gathered for each year during 1983–2002. Commodity data identified Coal Operator, Metal Operator, Nonmetal Operator, Stone Operator, Sand and Gravel Operator, Coal Contractor, and Noncoal Contractor categories. *Work location* referred to either underground or surface location. Injury data included the nonfatal days lost (NFDL) pertaining to permanent disability (partial or total), days away from work only, days away and restricted activity, and days of restricted activity only. *Employee hours* are the annual number of employee hours (N) exposed to risk of injury corresponding to a specific *commodity* and *work location*. The injury rates are calculated by dividing the NFDL by the corresponding *employee hours* and adjusted for 100 full-time employees or 200,000 employee hours. Fatality data is the number of fatalities during the year. *Employee hours* are the annual number of employee hours (N) exposed to the risk of fatality corresponding to a specific *commodity* and *work location*. The fatality rates are calculated by dividing the number of fatalities by the corresponding *employee hours* and adjusted for 100,000 full-time employees or 200,000,000 employee hours.

Table 1
Criteria for assessing goodness of fit of the Poisson model for non-fatal injuries

Criterion	DF	Value	Value/DF
Deviance	251	15,350.1163	61.1558
Scaled Deviance	251	15,350.1163	61.1558
Pearson Chi-Square	251	15,147.3217	60.3479
Scaled Pearson X2	251	15,147.3217	60.3479
Log Likelihood	2,071,261.1598		

Seven indicator variables are coded. These have a value of unity if the following are satisfied, otherwise their value is zero:

XC1=	Non Coal Contractor
XC2=	Metal Operator
XC3=	Non Metal Operator
XC4=	Stone Operator
XC5=	Sand and Gravel Operator
XC6=	Coal Contractor
XUGS=	Surface Work Location

Year corresponding to the data from 1983 to 2002 is represented by the continuous variable: RelYear=1 to 20 (1983 to 2002). Some researchers might prefer coding 0–19 for RelYear. This would only shift the constant term and related statistics but not the predictor variable parameter estimates including the one for RelYear.

2.2. Statistical analyses

The objectives of these analyses are to model the non-fatal and fatal injury rates in terms of the *commodity* and *work location* variables and to assess the injury rate for various *commodities* relative to the Coal Operator and the rate for *surface* work location relative to the *underground* work location. Another objective is to assess the annual reduction in injury rate during the selected 20 year period.

Table 2
SAS output for the NB model for non-fatal injuries

Model information

Data set	VJLIB.INJ_RAT1
Distribution	Negative binomial
Link function	Log
Dependent variable	Numinj
Offset variable	l_employ
Observations used	260

Criteria for assessing goodness of fit

Criterion	DF	Value	Value/DF
Deviance	251	266.4198	1.0614
Scaled deviance	251	266.4198	1.0614
Pearson chi-square	251	261.1170	1.0403
Scaled Pearson X2	251	261.1170	1.0403
Log likelihood	2,078,317.9399	Algorithm converged	

Analysis of parameter estimates

Parameter	DF	Standard estimate	Wald 95% error	Confidence	Limits	Chi-square	Pr>ChiSq
Intercept	1	−2.2419	0.0604	−2.3603	−2.1235	1377.39	<.0001
XC1	1	−0.3616	0.0678	−0.4945	−0.2288	28.46	<.0001
XC2	1	−0.3172	0.0652	−0.4449	−0.1895	23.70	<.0001
XC3	1	−0.4701	0.0661	−0.5997	−0.3405	50.55	<.0001
XC4	1	−0.2671	0.0693	−0.4029	−0.1312	14.85	<.0001
XC5	1	−0.1797	0.0828	−0.3420	−0.0175	4.72	0.0299
XC6	1	−0.0142	0.0702	−0.1518	0.1233	0.04	0.8393
XUGS	1	−0.7450	0.0396	−0.8225	−0.6674	354.55	<.0001
RelYear	1	−0.0170	0.0034	−0.0236	−0.0104	25.24	<.0001
Dispersion	1	0.0836	0.0093	0.0672	0.1041		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

Calculation of adjusted and unadjusted R2 measures:
number of observations=260
number of fitted df including intercept=10
R2 measures based on the log-likelihood:
unadjusted=0.6855665
adjusted with df=0.6742469
adjusted with method 1=0.6749445
adjusted with method 2=0.6741489

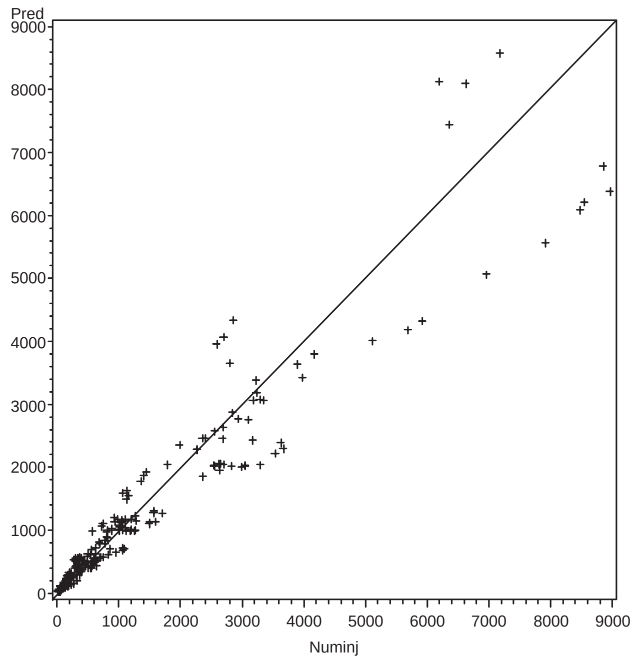

Fig. 1. Plot showing predicted versus actual non-fatal injury values.

Injury data are count or discrete event based and not continuous outcome based data. Multiple regression models may not be used for studying discrete data, if the data exhibits autocorrelation.

Based on the literature, both the Poisson and the negative binomial regression models were selected for studying their appropriateness to model the injury rate data. The following is a brief description of the two models.

Poisson regression model:

$$f(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \text{for} i = 0, 1, 2, \ldots$$

where *f(y_i)* is the probability distribution of y_i which is the number of injuries for the ith case and μ_i is the corresponding mean and $y_i!$ is the factorial of y_i.

The true underlying mean for the population is modeled in the log linear form:

$$\log(\mu) = \beta_0 + \beta_1 \text{XC1} + \ldots + \beta_6 \text{XC6} + \beta_7 \text{XUGS} + \beta_8 \text{RelYear} + \log(N)$$

where β is the regression coefficient relating the corresponding predictor variable to the mean number of injuries and log (N) is the offset link function for the generalized linear model.

Poisson regression model assumes that the mean and variance are equal. In the case of overdispersed data, the true variance is larger than the mean. For underdispersed data, the true variance is smaller than the mean. Evidence of overdispersion or underdispersion indicates inadequate fit of the Poisson model. A negative binomial model accounting for the overdispersion by including an error term in the regression model is appropriate for analyzing an overdispersed data. Yet, there are no published studies utilizing the negative binomial model for studying the injury rate data in mines and quarries.

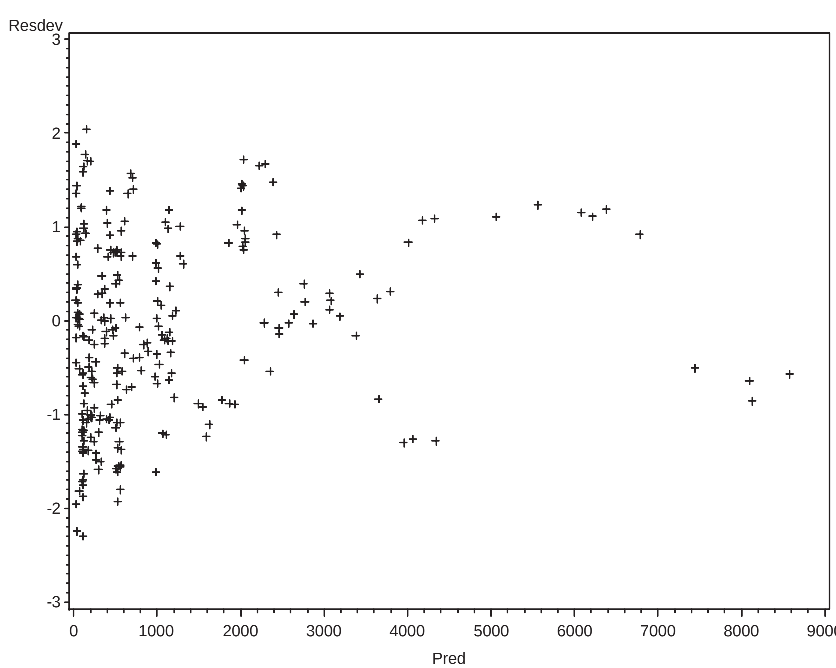

Fig. 2. Plot showing residual deviance for model based non-fatal injury predictions.

Negative binomial (NB) model:

$$f(y_i) = \frac{\Gamma(y_i + 1/\phi)}{\Gamma(y_i + 1)\Gamma(1/\phi)} \frac{(\phi\mu_i)^{y_i}}{(1+\phi\mu_i)^{(y_i+1/\phi)}} \text{for } i = 0, 1, 2\ldots$$

where $f(y_i)$, y_i, and μ_i are as explained above, ϕ is the dispersion parameter and Γ is the gamma function. Variance is given by:

$$\text{Var} = (Y) = \mu + \phi\mu^2.$$

PROC GENMOD in SAS version 8 (1999) is used to fit the Poisson and NB regression models. It employs the maximum likelihood approach for estimating the parameters.

In regression models the degree to which variation in the dependent variable can be explained by the predictor variables is also of interest besides the estimation of parameters and the significance of those variables. In multiple regression modeling the adjusted R^2 value is used for assessing the explained variance of the dependent variable as well as the effect of including additional predictor variables. For the Poisson regression model there are several similar approaches. Mittlböck (2002) provided a SAS-macro for calculating deviance based adjusted R^2 measures for the Poisson model, using the quasi likelihood function:

$$l_i = y_i \log(\mu_i) - \mu_i$$

For the negative binomial model, this author adapted the Mittlböck macro using the quasi likelihood function:

$$l_i = y_i \log(\phi\mu_i) - (y_i + 1/\phi)\log(1 + \phi\mu_i)$$

The corresponding SAS-macro is furnished in the Appendix.

3. Results and discussion

Table 1 shows the PROC GENMOD output for assessing goodness-of-fit of the Poisson model. The values of the

Table 3
SAS Output for the Poisson model for fatalities

Model Information

Data set	VJLIB.INJ_RAT1
Distribution	Poisson
Link function	Log
Dependent variable	Numinj
Offset variable	l_employ
Observations used	260

Criteria for assessing goodness of fit

Criterion	DF	Value	Value/DF
Deviance	251	400.2825	1.5948
Scaled deviance	251	400.2825	1.5948
Pearson chi-square	251	404.8111	1.6128
Scaled Pearson X2	251	404.8111	1.6128
Log likelihood	3392.6305	Algorithm converged.	

Analysis of parameter estimates

Parameter	DF	Standard estimate	Wald 95% error	Confidence	Limits	Chi-square	Pr>ChiSq
Intercept	1	−7.1196	0.0477	−7.2130	−7.0261	22,288.8	<.0001
XC1	1	1.2084	0.0836	1.0444	1.3723	208.70	<.0001
XC2	1	−0.0596	0.0799	−0.2162	0.0969	0.56	0.4551
XC3	1	−0.6417	0.1239	−0.8845	−0.3989	26.84	<.0001
XC4	1	0.0796	0.0745	−0.0665	0.2257	1.14	0.2859
XC5	1	0.4720	0.0883	0.2990	0.6450	28.60	<.0001
XC6	1	0.8309	0.0985	0.6379	1.0240	71.18	<.0001
XUGS	1	−1.0299	0.0561	−1.1399	−0.9199	336.75	<.0001
RelYear	1	−0.0322	0.0040	−0.0399	−0.0244	66.40	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Calculation of adjusted and unadjusted R2 measures:
number of observations=260
number of fitted df including intercept=9
R2 measures based on the log-likelihood:
unadjusted=0.6384209
adjusted with df=0.6268965
adjusted with method 1=0.6311944
adjusted with method 2=0.6306248

deviance and Pearson chi-square divided by the degrees of freedom (df) are used to detect overdispersion or underdispersion. Values greater than unity indicate overdispersion. This ratio at the indicated level of over 60 far exceeded the desired level of unity. The assumption then can be made that the injury counts are following a negative binomial model. The Poisson model is therefore dropped and the NB model is pursued for fitting the injury rate data.

Table 2 shows the PROC GENMOD output for the NB model. The ratio, value to degrees of freedom, is very close to unity showing the excellent goodness of fit of the negative binomial model to the injury data. Also, the dispersion parameter (0.0836) is positive indicating overdispersion in the data, and therefore, the appropriateness of the NB model over the Poisson model.

The relative effect of XC1 on injury rate is exp (−0.3616)=0.6966 (i.e., the mean injury rate for non coal contractor is 30.34% lower relative to the coal operator). Relative effects on injury rate for others are:

Metal Operator:	27.18% lower
Non Metal Operator:	37.51% lower
Stone Operator:	23.44% lower
Sand and Gravel Operator:	16.45% lower
Coal Contractor:	1.41% lower

Most interestingly, the relative effect of XUGS on the injury rate is exp (−0.7450)=0.4747 (i.e., the mean injury rate in surface mining is 52.53% lower relative to the rate for underground mining).

The negative sign for the RelYear parameter shows the decreasing trend of the injury rate over the 20 year period. The mean injury rate decreased by (1-exp (−0.017)=) 1.69% annually.

The parameter for XC6 (Pr=0.8393) is found to be insignificant. Though it could be dropped from the regression, thus creating a more complex Coal Operator/ Contractor as the basis, it was retained to maintain the easy to understand Coal Operator basis.

Fig. 1 shows a reasonably good fit of the actual versus predicted injuries data from the NB model. The line with a slope of unity depicting the perfect prediction of the actual injuries serves as reference for this assessment. The deviance based adjusted R^2 values are in excess of 67%, further indicating the reasonably good predictability of the NB model. These R^2 values measure the relative reduction in deviance due to predictors in the model.

Fig. 2 shows the plot of deviance residuals versus model based injury predictions. The pattern shows a satisfactory spread of the residuals.

Based on the model, the mean injuries can be predicted by the equation:

$$\mu = 0.1063(0.6966)^{\mathrm{XC1}}(0.7282)^{\mathrm{XC2}}(0.6249)^{\mathrm{XC3}} \times (0.7656)^{\mathrm{XC4}}(0.8355)^{\mathrm{XC5}}(0.9859)^{\mathrm{XC6}}(0.4747)^{\mathrm{XUGS}} \times (0.9831)^{\mathrm{RelYear}} N.$$

Table 3 shows the PROC GENMOD output from the Poisson model for the fatal injury data. The ratio, value to degrees of freedom, at the level of about 1.59 shows the reasonably good fit of the Poisson model to the fatality data.

Both the parameters for XC2 (Pr=0.4551), and XC4 (Pr=0.2859) are found to be less significant. Though these two variables could be dropped from the regression thus creating a more complex Coal/ Metal/Stone Operator basis, they were retained to maintain the easy to understand Coal Operator basis.

The relative effect of XC1 on fatality rate is exp (1.2084)=3.3481 (i.e., the mean fatality rate for Non-Coal contractor is 234.81% higher relative to the rate for the Coal Operator). Relative effects on fatality rate for others are:

Metal Operator:	5.79% lower
Non-Metal Operator:	47.36% lower
Stone Operator:	8.29% higher
Sand and Gravel Operator:	60.32% higher
Coal Contractor:	129.54% higher

The relative effect of XUGS on the fatality rate is exp (−1.0299)=0.3570 (i.e., the mean fatality rate in Surface Mining work location is 64.30% lower relative to the rate in Underground Mining work location).

The negative sign for the RelYear parameter shows the decreasing trend of the fatality rate over the 20 year period. The mean fatality rate decreased by (1-exp (−0.0322)=) 3.17% annually.

Fig. 3 shows a reasonably good fit of the actual versus predicted fatality data from the Poisson model. The line with a slope of unity depicting the perfect prediction of the actual

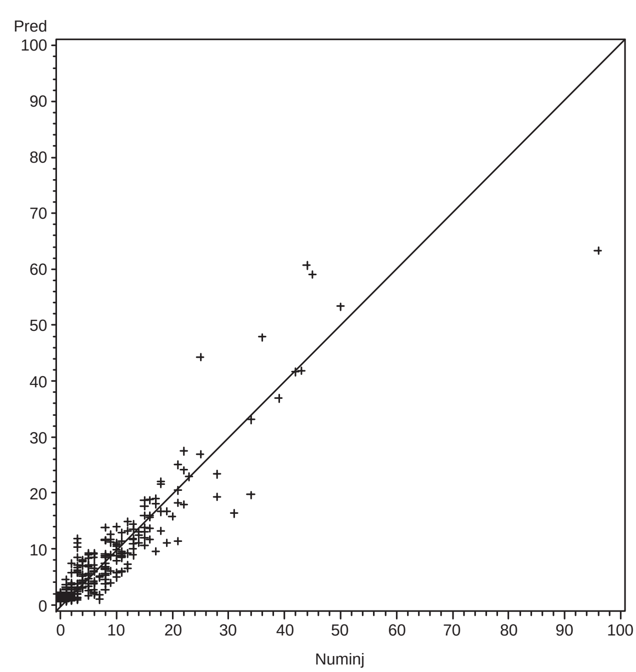

Fig. 3. Plot showing predicted versus actual fatality values.

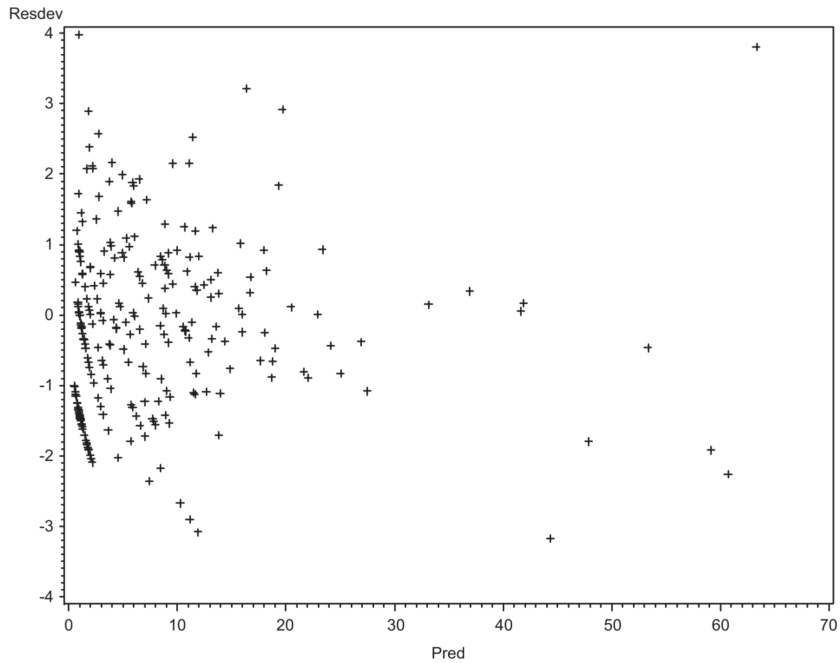

Fig. 4. Plot showing deviance residuals versus model based fatality predictions.

fatalities serves as reference for this assessment. The deviance based adjusted R^2 values are about 63%, indicating the satisfactory fatality predictions by the Poisson regression model. These R^2 values measure the relative reduction in deviance due to predictors in the model.

Fig. 4 shows a satisfactory pattern for the deviance based residuals.

Based on the model, the mean fatalities can be predicted by the equation:

$$\mu = 0.0008091(3.3481)^{XC1}(0.9421)^{XC2}(0.5264)^{XC3} \times (1.0829)^{XC4}(1.6032)^{XC5}(2.2954)^{XC6}(0.3570)^{XUGS} \times (0.9683)^{RelYear}N.$$

As with any regression model, one can only ascertain the relation between the dependent and the independent variables, but can never be certain of any underlying causal mechanisms. Extrapolation of the models to future time periods is inappropriate. Independent variables (such as age, experience of miners) not investigated in this study might have some significance in explaining the variation in the dependent variable.

Future work is to focus on a few key commodities and analyze the injury data for identifying the source and nature of those injuries. This is essential for developing interventions with significant potential for reducing injuries to miners.

4. Conclusions

The 20 year injury rate data for workers in the mining of various commodities can be adequately represented by the negative binomial model and the mean injuries rates, using the Coal Operator and Underground Work Location as the references, can be predicted satisfactorily from the relation:

$$(\mu/N) = 0.1063(0.6966)^{XC1}(0.7282)^{XC2}(0.6249)^{XC3} \times (0.7656)^{XC4}(0.8355)^{XC5}(0.9859)^{XC6} \times (0.4747)^{XUGS}(0.9831)^{RelYear}.$$

For the period 1983–2002, the annual reduction in the injury rate is 1.69%.

The 20 year fatality rate data for workers employed in the mining of various commodities can be adequately represented by the Poisson model and the mean fatality rates, using the Coal Operator and Underground Work Location as the references, can be predicted satisfactorily from the relation:

$$(\mu/N) = 0.0008091(3.3481)^{XC1}(0.9421)^{XC2}(0.5264)^{XC3} \times (1.0829)^{XC4}(1.6032)^{XC5}(2.2954)^{XC6} \times (0.3570)^{XUGS}(0.9683)^{RelYear}.$$

For the period 1983–2002, annual reduction in the fatality rate is 3.17%.

Acknowledgements

The author is grateful to the SSRSA branch of Pittsburgh Research Laboratory for mining of the non-fatal and fatal injury data from the MSHA database and to Dr. Martina Mittlböck, University of Vienna, Austria, in reviewing the R^2 relationship for assessing the explained deviance of the negative binomial model and for the SAS macro for adjusted R^2 measures at<http://www.akh-wien.ac.at/imc/biometrie/r2poi.htm>.

References

Bailer, A. J., Reed, L. D., & Stayner, L. T. (1997). Modeling Fatal Injury Rates Using Poisson Regression: A Case Study of Workers in Agriculture, Forestry, and Fishing. *Journal of Safety Research*, *28*(3), 177–186.

Bennet, J. D., & Passmore, D. L. (1984). Probability of Death, Disability, and Restricted Work Activity in United States Underground Bituminous Coal Mines, 1975–1981. *Journal of Safety Research*, *15*(2), 69–76.

Byers, A. L., Allore, H., Gill, T. H., & Peduzzi, P. N. (2003). Application of negative binomial modeling for discrete outcomes: A case study in aging Research. *Journal of Clinical Epidemiology*, *56*, 559–564.

Hull, B. P., Leigh, J., Driscoll, T. R., & Mandryk, J. (1996). Factors Associated with Occupational Injury Severity in the New South Wales Underground Coal Mining Industry. *Safety Science*, *21*, 191–204.

Kerkering, J.C., & McWilliams, P.C. (1987). Indices of Mine Safety Resulting from the Application of the Poisson Distribution to Mine Accident Data. United States Department of Interior, Bureau of Mines Report of Investigations, RI 9088, 31.

Khattak, A. J., Khattak, A. J., & Council, F. M. (2002). Effects of Work Zone Presence on Injury and Non-Injury Crashes. *Accident Analysis and Prevention*, *34*, 19–29.

Maiti, J. (2003). Technical Note: Development of Risk Indices for Underground Coal Mine Workers in India. *Mining Technology (Trans. Inst. Min. Metall. A)*, *112*, A119–A124.

Maiti, J., & Bhattacharya, A. (1999). Evaluation of Risk of Occupational Injuries Among Underground Coal Mine Workers Through Multinomial Logit Analysis. *Journal of Safety Research*, *30*(2), 93–101.

Miaou, S. (1994). The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions. *Accident Analysis and Prevention*, *26*(4), 471–482.

Mittlböck, M. (2002). Calculating Adjusted R^2 Measures for Poisson Regression Models. *Computer Methods and Programs in Biomedicine*, *68*, 205–214.

National Institute for Occupational Safety and Health [NIOSH]. (2004). *1983–2003 Occupational Mining Fatality and Injury Rates by Commodity and Underground/ Surface Work Location,* http://pit/prl/surveillance/tables/r11.html

SAS. (1999). Cary, NC: SAS Institute Inc.

Appendix A. Mittlböck SAS macro for adjusted R^2 adapted for the NB model

```
MACRO R2POI (DATA=, DEP=, TOTAL=, CLASS= , CONT=  );

data &data; set &data;
 %if &total=  %then %do;
  _total=1;
  %let total=_total;
 %end;
 run;
ods output ObStats=pred ParameterEstimates=pest;
ods exclude listing obstats;
proc genmod data=vjlib.inj_rat1;
    class commod ugorsurf;
          model &dep= &class &cont / dist=nb link=log
    offset=l_employ obstats maxiter=100;
 run;
proc iml;
 file print;
 use &data var{&total};
  read all into total var{&total};
 close &data;
 use pest var{Parameter DF Estimate};
  read all into Estimate var{Estimate};
        read all into var var{DF};
        read all into Parameter var{Parameter};
 close pest;
 use pred var{&dep pred};
  read all into y var{&dep} where(pred^=.);
  read all into p var{pred} where(pred^=.);
 close pred;
 n=nrow(p);
 k=sum(var);
```

```
 mu  bar=sum(p)/sum(total);
 put;
 put "Calculation of adjusted and unadjusted R2 measures:";
 put "=================================================";
 put;
 put "number of observations = " n ;
 put "number of fitted df including intercept = " k ;
*****************************************************************;
** R2 based on log-likelihood for Negative Binomial model     **;
*****************************************************************;
 phi=Estimate[loc(parameter='Dispersion'),];
 d_sat=y[loc(y)];
 d_sat=sum(d_sat#log(phi#d_sat)-(d_sat+(1/phi))#log(1+phi#d_sat));
 d_0=sum(y#log(phi#mu_bar#total) - (y+(1/phi))#
 log(1+phi#mu_bar#total));
 d_b=sum(y#log(phi#p)-(y+(1/phi))# log(1+phi#p));
 d0=d_sat-d_0;
 dbeta=d_sat-d_b;
 r2=1-dbeta/d0;
 r2_df=1-(dbeta/(n-k))/(d0/(n-1));
 r2_e1=1-(dbeta+(k-1)/2)/d0;
 r2_e2=1-(dbeta+k/2)/(d0+0.5);
 put;
 put "R2 measures based on the log-likelihood:";
 put "-----------------------------------------";
 put "  unadjusted            = " r2 ;
 if r2_df<0 then do;
  put "  adjusted with df       = 0 (a negative value of " r2_df " was calculated)" ;
 end;
 else do;
  put "  adjusted with df       = " r2_df ;
 end;
 if r2_e1<0 then do;
  put "  adjusted with method 1 = 0 (a negative value of " r2_e1 " was calculated)" ;
 end;
 else do;
  put "  adjusted with method 1 = " r2_e1 ;
 end;
 if r2_e2<0 then do;
  put "  adjusted with method 2 = 0 (a negative value of " r2_e2 " was calculated)" ;
 end;
 else do;
  put "  adjusted with method 2 = " r2_e2 ;
 end;
quit;
%MEND

%R2POI(data=vjlib.sasvjtest4, dep=numinj, total=emphrx, class=xc1 xc2 xc3 xc4 xc5 xc6 xugs,
cont=RelYear);
quit;
```

Vijia Karra received MS and PhD in Materials Science and Engineering from the University of California, and MBA from the University of Wisconsin. Currently, he is Chief of the Mining Injury Prevention Branch, Pittsburgh Research Laboratory, National Institute for Occupational Safety and Health. Prior to this, he was a Vice President in Technology/Business Development at Metso Minerals. He is a member of the Society of Mining Engineers and has over 100 publications and patents to his credit.