THE USE OF A MIXED WEIBULL MODEL IN OCCUPATIONAL INJURY ANALYSIS

MIN K. CHUNG

Department of Industrial and Systems Engineering, The University of Illinois at Chicago, Chicago, IL 60680 (U.S.A.)

SHU-CHEN H. WU

Division of Epidemiology and Health Computer Sciences, The University of Minnesota, Minneapolis, MN 55455 (U.S.A.)

GARY D. HERRIN

Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109 (U.S.A.)

ABSTRACT

Chung, M.K., Wu, S-C.H. and Herrin, G.D. (1986). The use of a mixed Weibull model in occupational injury analysis. *Journal of Occupational Accidents*, 7: 239–250.

This article describes a mixed Weibull model which is proposed as an alternative model of occupational injury analysis. This model assumes that a worker will suffer injuries during employment with a probability γ ($0 < \gamma < 1$). The conditional failure time model is defined to be a Weibull distribution. Given the model, the relationship between minor and major injury incidents is examined using the injury data of 1,004 workers in a southwestern industrial plant. The efficacy of the model is apparent in that it provides a fraction of long-term survivors as well as injury rate for those ever suffering an injury.

1. INTRODUCTION

The use of sound statistical techniques is very important in identifying agents causing injuries and illnesses in the work environment. Common methods of computing, summarizing and comparing these accident experiences have been developed and widely used. Such methods include the computation of incidence rate and severity rate (e.g., American National Standards Institute, 1973), and the use of contingency tables to test an association between variables of interest (Fragala, 1983). The incidence rate, however, is less reliable because minor injury incidents are often ignored or under-reported (Tarrants, 1980). Even for major injuries such as those required to be reported by the Occupational Safety and Health Act (OSHA, 1970), the aggregate statistics do not adequately reflect relative injury risks

due in part to the rarity of these incidents. Minor injuries, in this context, are those that require some type of medical treatment, but do not result in death, lost workdays or transfer to another job. Major injuries, however, are associated with death, lost workdays, or job reassignment. Contingency tables are easy to use, but are not adequate for recurring events such as injury. Furthermore, the computation of injury rate assumes that every incident is equally likely to occur to every individual. The validity of this assumption is questionable because different types of workers and work environments are often interspersed within one department.

As an alternative model of injury analysis, a mixed Weibull model is proposed in this paper. This model assumes that some fraction of the worker population will not be injured. This is in contrast to conventional methods which assume an equal risk of injury for all workers. The conditional failure time model is defined to be a Weibull distribution.

The purpose of this paper is to construct an alternative model of occupational injury analysis based on more realistic assumptions, and to examine the fidelity of the model using the actual injury data. The injury data for 1,004 workers have been provided by a large manufacturing company located in the southwestern United States. The employees' work history and medical records from 1 January 1980 through 31 December 1984 are maintained in a computer database, called the Occupational Health and Safety Information Management System (OHSIMS), developed at the University of Michigan Center for Ergonomics (Anderson and Chung, 1983).

Given the model, the following hypotheses are examined: (1) workers' minor injury statistics can be used as predictions of subsequent major injury; (2) workers' injury recidivism or injury proneness can be determined by comparing the first major injury experiences with the second major injury experiences; (3) worker/job matching in terms of gender and physical stress is associated with the risk of suffering a major injury.

The parameters of the Weibull model and the probability of ever getting injured are estimated for the first and second major injury data, respectively. To tackle the hypotheses (1) and (3) stated above, covariate analysis is performed by making assumptions that the scale parameter of the Weibull model is a log-linear function of covariates, and the probability of getting injured is a logistic model of covariates (Farewell, 1977). The analyses are carried out for high and low job stress categories separately.

2. MODEL DEVELOPMENT

The injury behavior of workers is examined using a mixture model for failure time analysis. This model assumes that a worker will suffer a major injury during his/her employment with a probability γ ($0 < \gamma < 1$). The conditional failure time model is assumed to be a Weibull distribution. Then the parameters of the mixture model provide information about two useful quantities: (i) a fraction of long-term survivors (i.e., those who never suffer a

major injury during employment), and (ii) the injury rate for those ever suffering a major injury. We examine the first and second major injury incidents separately.

2.1 Defining the failure event

The first step in the modeling process is to define the failure event. The definition of the failure event depends upon the application. For example, in machine failure applications, the failure event may be defined in terms of components' wear (Kapur and Lamberson, 1977). For occupational injury analysis, there may be a number of ways of defining the failure event. It may be any recordable injury or illness defined by the Occupational Safety and Health Act; any visit to a dispensary unit for job-related incidents; or any disabling injury. In this study, the failure event is defined as any occupational injury incident that resulted in one or more lost workdays. Such an incident is called a major injury. The first major injury time is defined as the length of time an employee worked on a job continuously until he experienced the first major injury. If an employee had no major injury during the study period or he left the job without having any major injury before the end of the study, this is treated as a censored observation. The second major injury time is defined as the length of time between recovery from the first major injury and experiencing the second major injury. The number of days from the date of the first injury until returning to the job is defined as lost days or restricted days attributable to the first injury.

2.2 Notation and model

A binary variable X is defined as follows:

$$X = \begin{cases} 0 & \text{if a worker will not suffer a major injury while performing a job} \\ 1 & \text{if a worker will suffer a major injury while performing a job} \end{cases}$$

Then $\gamma = \Pr(X = 1)$ is defined as the probability that a worker will suffer a major injury. The injury times, t, of workers for whom $X = 1$, whether first injury or second injury, are described using a 2-parameter Weibull distribution.

The probability density function of the Weibull distribution is expressed as

$$f(t \mid X = 1) = \delta \lambda (\lambda t)^{\delta - 1} \exp [-(\lambda t)^{\delta}], \quad t > 0 \tag{1}$$

where δ and λ are shape and scale parameters respectively. The injury rate (or hazard rate) can be written as

$$h(t \mid X = 1) = \delta \lambda (\lambda t)^{\delta - 1}, \quad t > 0 \tag{2}$$

which is the instantaneous injury rate at time t given that an individual has

not suffered any major injury prior to time t. Hence the injury rate is monotonically increasing over time if $\delta > 1$ and monotonically decreasing if $0 < \delta < 1$. If $\delta = 1$, the injury rate is constant, in which case the distribution becomes an exponential type.

2.3 Parameter estimation

The parameters of the injury time model described above are estimated by a maximum likelihood technique. The ith worker who suffered a major injury at time t_i contributes a likelihood factor, L_i, which is the probability of suffering an injury at time t_i. It can be expressed as

$$L_i = Pr(X_i = 1)\, f(t_i|X_i = 1) = \gamma\delta\lambda(\lambda t_i)^{\delta - 1}\, \exp\left[-(\lambda t_i)^\delta\right] \tag{3}$$

The jth worker who had been followed to time t_j without suffering any injury contributes a likelihood factor, L_j, which is the sum of the probability of being a long-term survivor and the probability of suffering an injury that occurs after time t_j. It can be expressed as

$$L_j = Pr(X_j = 0) + Pr(X_j = 1) \int_t^\infty f(u|X_j = 1)\, du = 1 - \gamma + \gamma \exp\left[-(\lambda t_j)^\delta\right] \tag{4}$$

Suppose that k employees suffered injuries and $n - k$ employees survived without suffering any injury during the study period. The likelihood function can then be constructed as follows:

$$L(\gamma,\delta,\lambda) = \prod_{i=1}^{k} L_i \prod_{j=k+1}^{n} \tag{5}$$

The maximum likelihood estimates (M.L.E.) can be obtained by setting equal to zero the partial derivatives of the logarithm of the likelihood function (5) with respect to γ, δ and λ. It is difficult analytically to find M.L.E.'s for γ, δ, and λ. Hence a numerical non-linear minimization technique based on quasi-Newton methods is used.

The asymptotic distribution of the maximum likelihood estimators, $\hat{\underset{\sim}{\theta}} = (\hat{\gamma},\hat{\delta},\hat{\lambda})$, is, under certain regularity conditions, multivariate normal with mean $\underset{\sim}{\theta}$ and variance-covariance matrix $(\underset{\sim}{\theta}) = I(\underset{\sim}{\theta})^{-1}$ (Cox and Hinkley, 1974). An estimator of $I(\underset{\sim}{\theta})$ can be provided by the observed information matrix:

$$\hat{I}(\underset{\sim}{\theta}) = -\left[\frac{d^2 \ln L(\underset{\sim}{\theta})}{d\theta_i d\theta_j}\bigg|\, \underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}\right] \tag{6}$$

2.4 Covariate analysis

The sample population can also be characterized in terms of specific factors such as race, gender, age, geographic location or other characteristics. Some, if not all, of these factors may be expected to have a relationship with injury (failure) rates. In failure time analysis, the variables describing these factors are called concomitant or explanatory variables or, briefly, covariates (Elandt-Johnson and Johnson, 1980).

Covariate values of each individual are also taken into account by defining the parameters of the injury time model as a function of the covariates. A logistic model is used for the probability parameter, γ, which is associated with the binary variable X defined above. A log-linear model is used for the scale parameter, λ, of the Weibull distribution (Farewell, 1977). However, the shape parameter, δ, of the Weibull distribution is assumed to be independent of the covariate values.

Two covariates are considered for each worker in this study: (i) gender and (ii) minor injury experience. A worker's minor injury experience is measured in terms of the incidence rate for minor injury incidents per year. The minor injury incidence rate (IR) is computed as follows:

$$IR = (M/H) \times 2000 \tag{7}$$

where M = number of minor injuries incurred,

H = hours worked prior to the first major injury incident if any major injury was incurred, or hours worked until the end of the study if no major injury was incurred, and

2000 = base for a full-time equivalent employee working 40 hours per week, 50 weeks per year.

Hence the probability parameter γ is defined as follows:

$$\gamma(\underset{\sim}{z}) = Pr(x = 1 | \underset{\sim}{z}) = \exp(\underset{\sim}{v}\underset{\sim}{z})/(1 + \exp(\underset{\sim}{v}\underset{\sim}{z})) \tag{8}$$

where $\underset{\sim}{z} = (1, z_1, z_2)$ is the covariate vector, $z_1 = 0$ if male and $z_1 = 1$ if female, and z_2 is the minor injury incidence rate per year, and $\underset{\sim}{v} = (v_0, v_1, v_2)$ is the corresponding parameter vector. The scale parameter, λ, of the Weibull distribution is also defined as follows:

$$\lambda(\underset{\sim}{z}) = \exp(\underset{\sim}{w}\underset{\sim}{z}) \tag{9}$$

where $\underset{\sim}{w} = (w_0, w_1, w_2)$ is the parameter vector.

Although other models can be used, these models are chosen to keep the domain of these parameters such that $0 < \gamma(\underset{\sim}{z}) < 1$ and $\lambda(\underset{\sim}{z}) > 0$. The maximum likelihood estimates are again obtained for the parameter vector $\underset{\sim}{\theta}$ by substituting (8) and (9) into the likelihood functions, where $\underset{\sim}{\theta} = (v_0, v_1, v_2, \delta, w_0, w_1, w_2)$.

The variances and covariances of the parameters can be estimated from an observed information matrix discussed above. A normal deviate is used to test whether an individual covariate is a significant factor.

2.5 Injury recidivism/injury proneness

In this paper, the workers' injury recidivism or injury proneness is also examined. The term "recidivism" is often used in the studies of correctional programs. Maltz (1980) defined it as follows:

> A person is considered to have recidivated if his (criminal) behavior has not been improved by the correctional intervention. The work is derived from the Latin "recidere", to fall back. Thus a recidivist is one who, after his release from custody for having committed a crime, is not rehabilitated. Instead, he falls back, or relapses, into his former behavioral patterns: he commits more crimes.

In this study injury recidivism is used in the context of injury prevention. A relevant question addressed is whether a worker who had experienced the first major injury was "rehabilitated" with respect to the second major injury. Injury proneness is also examined by considering the question whether a worker who once suffered a major injury would have a greater likelihood of experiencing another major injury.

3. MODEL ILLUSTRATION

The model discussed above is illustrated using the injury data for 1,004 workers in a southwestern industrial plant. First, the relation between minor and major injury incidents is examined using the mixed Weibull model without considering covariates for each of the first and second major injury experiences. Workers' injury recidivism or injury proneness is also studied. Secondly, gender and minor injury experience are modeled as covariates, and their effects on the risk of suffering a major injury are investigated.

3.1 The relation between minor and major injuries

The population is divided into two groups, based on the minor injury incidence rate, in order to examine the relation between minor and major injuries. Workers who have experienced an average of one or less minor injury incidents per year are considered to be in a low minor injury group, otherwise they are in a high minor injury group. Table 1 shows the number of workers who experienced the first and second major injuries stratified by minor injury incidence rate. In the high minor injury incidence rate group, among 347 workers who suffered a major injury there were 277 of them who suffered a second major injury and 70 workers had not done so by the time the data were collected. The average lengths of duration before the first major injury occurred were 35.6, and 46.7 months in the high and low minor injury incidence rate groups, respectively. The average periods of time to second major injury from recovering from the first major injury were 19.3, and 28.2 months in the two groups, respectively.

The exact failure time data of 347 employees who experienced high minor injury incidents are used to test the fitness of the Weibull distribution (see

TABLE 1

Number of workers who experienced the first and second major injury

Minor injury* incidence rate	Major injury incident			
	First		Second	
	Injured	Not injured	Injured	Not injured
High	347	56	277	70
Low	254	347	169	85
Total	601	403	466	155

*Workers who have experienced an average of one or less minor injury incident per year are considered to be in a low minor injury group, otherwise they are in a high minor injury group.

Table 1). $F(t_{(i)})$ is estimated for the ith ordered failure time, $t_{(i)}$, and then these values are plotted on Weibull probability paper. Figure 1 shows the fitted line and plotted data points for $\ln t$ and $\ln [\ln [1/(1 - F(t))]]$ on Weibull probability paper. This plot depicts a straight line with approximate slope 1.17, although there is a minor departure from the straight line. The shape parameter δ is estimated graphically as the reciprocal of the slope of a straight line fitted to the data on the Weibull probability paper. The value for λ found from the 63.2 percentile is approximately 0.028. This supports the validity of the Weibull distribution assumption for injury times. These estimated values for δ and λ are used as initial values for parameter estimation. Although the graphically estimated value for δ, 0.85, is close to 1.0, suggesting an exponential approximation, it should be noted that censored observations are not considered in this estimation.

Table 2 displays the maximum likelihood estimates and 95% confidence intervals for the probability of suffering a major injury, γ, of the mixed Weibull model stratified by minor injury incidence rate. For both the first and second major injury experiences, the high minor injury group is consistently found to have a significantly greater probability of suffering a major injury than the low minor injury group ($p < 0.01$).

Figure 2 displays joint confidence regions for the shape and scale parameters δ and λ, while the probability parameter γ is held constant. In Fig. 3, the instantaneous injury rates for two groups are plotted for the first and second major injury experiences, respectively. The injury rates, for the first major injury experiences, of the high minor injury group also appear to be higher than those of the low minor injury group. However, after 6 or 7 years of exposure, they exhibit the opposite results due to larger values of the shape parameter for the low minor injury group as seen in Fig. 2. For comparison, the maximum likelihood estimates of the constant injury rate are also obtained using the conventional assumptions: 0.0242 and 0.0091 for
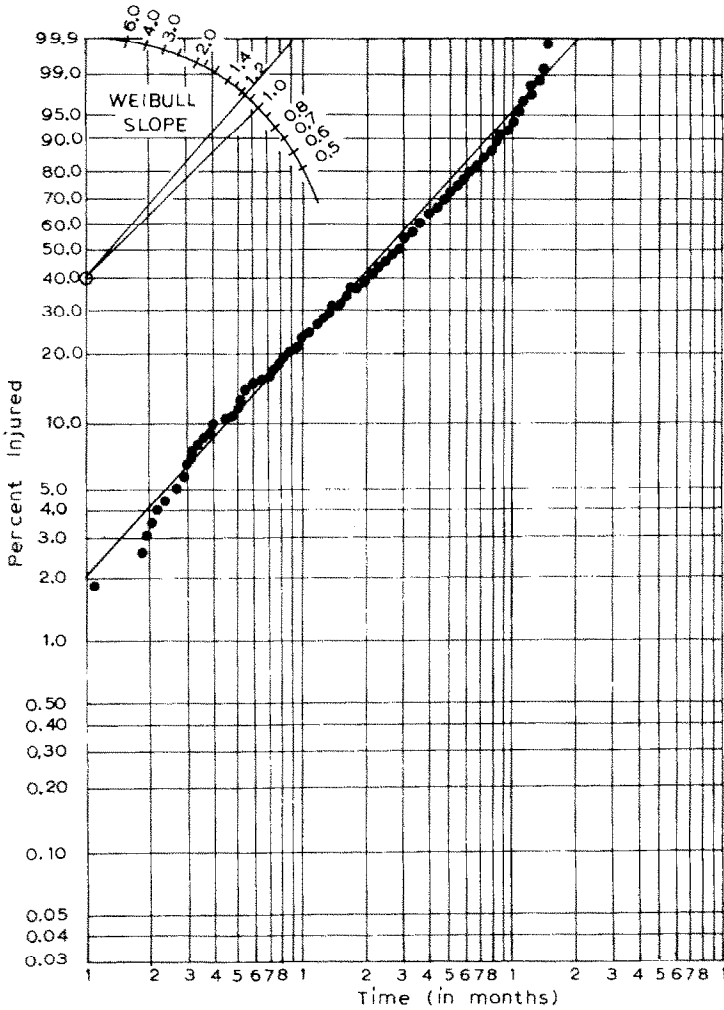
Fig. 1. The fitted line and plotted data points for ln [ln [1/(1 − F(t))]] and ln t on the Weibull probability paper.

TABLE 2

The maximum likelihood estimates and 95% confidence intervals for the parameter, γ, of the mixed Weibull model

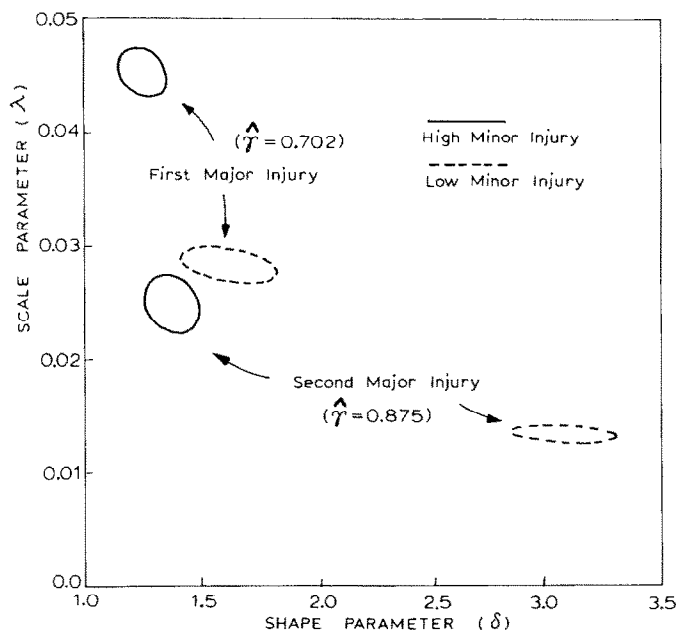| Minor injury incidence rate	First major injury		Second major injury	
	γ	CI[a]	γ	CI[a]
High	0.954	(0.918, 0.975)	0.945	(0.922, 0.962)
Low	0.702	(0.654, 0.745)	0.875	(0.818, 0.915)

[a] CI = 95% Confidence Interval.

Fig. 2. 95% confidence regions for the shape and scale parameters of Weibull distribution.
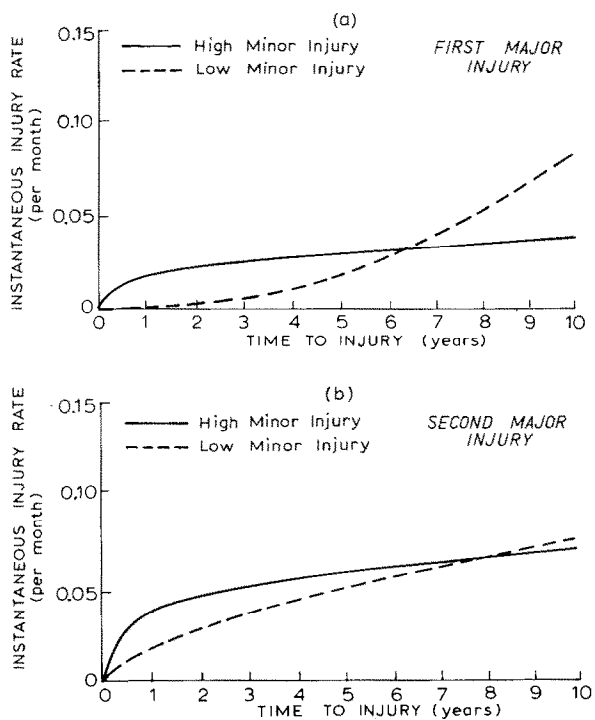

Fig. 3. Comparison of injury rates between high and low minor injury groups for (a) the first and (b) second major injuries.

high and low minor injury groups, respectively, for the first major injury experiences; and 0.0405 and 0.0236 for the high and low minor injury groups, respectively, for the second major injury experiences. These constant injury rates appear to be underestimated compared to those of the mixed Weibull model.

The results of the first and second major injury are compared in order to examine the workers' tendency of injury recidivism or injury proneness. For the high minor injury group, the injury rates for the second major injury appear to be higher than those for the first major injury; whereas there is no significant difference in the probability of suffering a major injury between the first and second major injuries ($p = 0.34$). For the low minor injury group, the injury rates for the second major injury also appear to be higher than those for the first major injury, although the latter increases faster due to larger values of the shape parameter (Fig. 2). The probability of suffering the second major injury is significantly greater than the probability of suffering the first major injury ($p < 0.01$). Thus both groups show the tendency of injury recidivism and injury proneness.

3.2 Covariate analysis

The effects of gender and minor injury experience upon the risk of suffering a major injury are examined using the covariate analysis method discussed in section 2.4. In the plant under study, physical stress is also a major concern. As a result, the worker population is divided into two groups depending on the physical stress of the job they are assigned: (i) high stress, and (ii) low stress job group. The criteria for high or low stress jobs are determined based on the results of the Biomechanical Strength Prediction Model developed at the University of Michigan's Center for Ergonomics (Garg and Chaffin, 1975). Then the analysis is performed separately for these two worker populations and is confined to the injury time to the first major injury.

Table 3 shows the number of workers stratified by gender and job stress type. One normally would expect that a smaller fraction of female workers would be assigned to high stress jobs as compared to male workers.

TABLE 3

Distribution of workers stratified by gender and job stress type

Job stresses	Male	Female	Total
High	548	75	623
Low	324	57	381
Total	872	132	1004

TABLE 4

Results of the maximum likelihood estimation for covariate analysis with respect to the first major injury time

Covariates	γ			λ		
	$\hat{v}$	s.e. $(\hat{v})$	normal deviate	$\hat{w}$	s.e. $(\hat{w})$	normal deviate
(a) High stress job group						
z_0	0.579	0.185		−4.168	0.039	
Gender	2.348	3.831	0.61	0.338	0.084	4.02[a]
Minor injury	5.450	1.119	4.87[a]	0.678	0.047	14.43[a]
(b) Low stress job group						
z_0	−0.793	0.256		−4.307	0.060	
Gender	0.365	0.821	0.44	0.113	0.106	1.07
Minor injury	6.433	1.687	3.81[a]	1.144	0.109	10.49[a]

[a] Significant at $\alpha = 0.001$
$\hat{\delta} = 1.94$

Surprisingly, this data exhibits almost the same fraction of female workers assigned to high stress jobs $(75/132 = 56.8\%)$ as male workers $(548/872 = 62.8\%)$, which suggests a possibility of worker/job mismatching.

Table 4 displays the maximum likelihood estimates for the parameters of the model with covariates and their normal deviate values among workers exposed to high and low stress jobs, respectively. For both groups, the effect of minor injury experience upon the scale parameter, λ, of the Weibull distribution is found to be highly significant $(p < 0.001)$. The effect of gender upon the scale parameter is highly significant $(p < 0.001)$ for the high stress job group by yielding a higher λ value for female workers, but it is not significant $(p = 0.29)$ for the low stress job group. The effect of minor injury experience upon the probability of suffering a major injury, γ, is highly significant $(p < 0.001)$ for both groups. There is no evidence of the association of gender and the probability of suffering a major injury for either group. The estimates for the shape parameters, δ, happen to be the same for both high and low stress job groups. This $\hat{\delta}$ value of 1.94, which is greater than 1, implies an increasing failure rate as shown in Fig. 3.

4. DISCUSSION

In conventional methods based on incidence rates, it is assumed that the entire worker population is subject to a constant injury rate. However, the proposed model assumes that a fraction of the worker population will not be injured, and that the injury rate increases or decreases over time. The efficacy of the latter model is apparent in that it provides a fraction of long-

term survivors as well as injury rate for those ever suffering an injury. These two quantities are also useful in examining workers' injury recidivism or injury proneness tendency.

As discussed above, the constant injury rates obtained by the conventional methods appear to be underestimated relative to those of the mixed Weibull model. This is due to the assumption of the injury rate being constant, and also due to the assumption that every worker will get injured with probability one. As Farewell (1982) pointed out, it should be noted that the assumption of a mixed population is too strong for general applications. However, in this study, the appropriateness of the mixture model application is supported by the evidence that there is a significant fraction of "no injury" workers in the plant under study.

The investigation of the effects of gender and minor injury experiences upon the risk of major injury leads to interesting and perhaps provocative results. Workers who have experienced frequent minor injury incidents show a greater probability of suffering a major injury than those who have experienced less frequent minor injury incidents. This result is very promising because it suggests the predictors of using minor injury statistics as predictions of subsequent major injury. It is also noted that the effect of gender upon the scale parameter of the Weibull distribution is highly significant for the high stress job group but is not significant for the low stress job group. This underscores the importance of a worker/job matching program that is based on a strength test or other tests related to physical stress of the job.

REFERENCES

American National Standards Institute, 1967. Methods of Recording and Measuring Work Injury Experience, ANSI Z16.1 (r. 1973).

Anderson, C.K. and Chung, M.K., 1983. "User's Guide to the Occupational Health and Safety Information Management Systems." Center for Ergonomics, The University of Michigan, September.

Cox, D.R. and Hinkley, D.V., 1974. Theoretical Statistics, Chapman and Hall, London.

Elandt-Johnson and Johnson, 1980. Survival Models and Data Analysis, John Wiley and Sons, Chichester.

Farewell, V.T., 1977. A model for a binary variable with time-censored observations. Biometrika, 64: 43—46.

Farewell, V.T., 1982. The use of mixture models for the analysis of survival data with long-term survivors. Biometrics, 38: 1041—1046.

Fragala, G., 1983. A modern approach to injury record keeping. Professional Safety, 28: 21—24.

Garg, A. and Chaffin, D.B., 1975. A biomechanical computerized simulation of human strength. AIIE Transaction, 7: 1—15.

Kapur, K.C. and Lamberson, L.R., 1977. Reliability in Engineering Design. John Wiley and Sons, Chichester.

Maltz, M.D., 1980. On Recidivism: Exploring Its Properties as a Measure of Correctional Effectiveness. Volumes I and II. Center for Research in Criminal Justice. University of Illinois at Chicago Circle.

Tarrants, W.E., 1980. The Measurement of Safety Performance, Garland, New York.