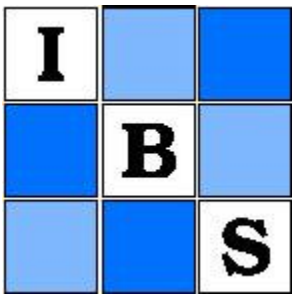

BIOMETRICS 48, 951–959
September 1992

Maximum Likelihood Estimation for Interval-Censored Data Using a Weibull-Based Accelerated Failure Time Model

Patricia M. Odell
Bryant College, Smithfield, Rhode Island 02917, U.S.A.
Keaven M. Anderson*
Centocor, Inc., Malvern, Pennsylvania 19355, U.S.A.
and
Ralph B. D'Agostino
Boston University, 111 Cummington Street, Boston, Massachusetts 02215, U.S.A.

SUMMARY

The accelerated failure time regression model is most commonly used with right-censored survival data. This report studies the use of a Weibull-based accelerated failure time regression model when left- and interval-censored data are also observed. Two alternative methods of analysis are considered. First, the maximum likelihood estimates (MLEs) for the observed censoring pattern are computed. These are compared with estimates where midpoints are substituted for left- and interval-censored data (midpoint estimator, or MDE). Simulation studies indicate that for relatively large samples there are many instances when the MLE is superior to the MDE. For samples where the hazard rate is flat or nearly so, or where the percentage of interval-censored data is small, the MDE is adequate. An example using Framingham Heart Study data is discussed.

1. Introduction

In epidemiological studies, it is possible to obtain exact occurrence times for most events. But there may be instances where times of occurrence of events are missing and can be established only to within specified intervals. For example, in the Framingham Heart Study, times of coronary heart disease (CHD) are usually readily available, but times of first occurrence of the CHD subcategory angina pectoris (AP) may be specified only as between two given clinic examinations, several years apart.

In this situation, time to event is said to be interval-censored. Interval censoring in fact includes the more familiar left- and right-censoring as extreme cases. In these instances, 0 or ∞ replaces the left or right endpoint of the interval.

For the usual situation of right-censoring, regression methods for determining the dependence of failure time T on covariates $X_1, X_2, \ldots, X_k$ are widely available, both for parametric and nonparametric models. [See, for example, Lawless (1982) and Kalbfleisch and Prentice (1980).] Relatively little has been published for the case of interval censoring, particularly where a parametric distribution is assumed. In analyses of the Framingham Study, interval midpoints have been substituted for the unknown event times, but little research has been done on the effects of such substitutions on estimation of parameters.

---

* Formerly with the National Heart, Lung, and Blood Institute in Framingham, Massachusetts.

Key words: Accelerated failure time model; Interval censoring; Weibull model.

Several studies have focused on interval censoring, but most deal with nonparametric or semiparametric models and procedures that have little applicability to the parametric case. For example, Finkelstein and Wolfe (1985) propose a semiparametric model for interval-censored data where no parametric model is assumed for T, but the conditional distribution of X given T follows a specified parametric model. Finkelstein (1986) presents a proportional hazards model, and Self and Grossman (1986) discuss linear rank statistics for interval-censored data.

There are two studies that do employ parametric models with interval censoring. Flygare, Austin, and Buckwater (1985) use a two-parameter Weibull distribution, not dependent on covariates, and compare estimates of the shape parameter where interval-based maximum likelihood estimates and midpoints, respectively, are substituted for missing event times. In most instances, the maximum likelihood estimates are found to be superior. In that study, all data are observed in intervals; i.e., all exact times are missing.

Brookmeyer and Goedert (1989) studied a cohort of hemophiliacs to determine the effects of covariates on infection with AIDS and progression to clinical disease once infected. The data are interval-censored, since exact times of events cannot be ascertained. The authors use a two-stage parametric (exponential/Weibull) model, in effect substituting interval-based maximum likelihood estimates for missing event times. No comparisons with other methods are made.

In this paper, we use simulations to assess the relative merits of interval-based maximum likelihood estimates and midpoint-based estimates of event times in the estimation of parameters. It might be expected that the latter would not do well for distributions where the midpoint is a poor estimate of the actual data. We define our procedure in terms of a Weibull-based accelerated failure time model because we have found this model to be very satisfactory in working with Framingham Study data (see Anderson, 1991). All censoring in simulations is random and independent of time to event. Simulation for other base models, such as the log-logistic or log-normal, could be accomplished using the same procedures.

In addition to the use of midpoints or maximum likelihood estimates, there are other possible methods for dealing with interval-censored event times. Imputation using an EM algorithm is one such option, but this procedure is not as straightforward as usual because the Weibull distribution is not a member of the exponential family. The maximum likelihood method is no more difficult to implement than most alternatives and offers theoretical and speed-of-convergence advantages. The substitution of interval midpoints ranks first in convenience, since parameter estimates may be obtained using standard software packages. Hence, the scope of this paper will extend only to a comparison of these two procedures. The example given in Section 3 uses Framingham Offspring Study data for angina pectoris, a condition whose hazard rate increases with age. The progression from infection to disease in AIDS would be another such example. An example of an event with decreasing hazard rate might be time to relapse after treatment of some types of cancer.

To accommodate a wide range of practical applications, we present the results of simulations using different shape parameters (from rapidly increasing to rapidly decreasing failure rates) for samples of various sizes and events with different rates of incidence. Based on these results, we conclude with case-specific recommendations as to which estimate is most useful.

Although we used a program in C, written by one of the authors, to generate data, run simulations, and do the examples, it is now quite straightforward to find maximum likelihood estimates for interval-censored data using SAS PROC NLIN or PROC LIFEREG (SAS, 1990).

2. The Model

The Weibull density function is written

$$f_T(t) = \alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}},$$

where α is a shape parameter and λ is a scale parameter such that $\lambda = e^{-\boldsymbol{\beta}'\mathbf{X}}$ where $\mathbf{X}$ is a vector of covariates $(X_1, X_2, \ldots, X_k)'$ and $\boldsymbol{\beta}$ is a vector of parameters $(\beta_1, \beta_2, \ldots, \beta_k)'$.

In this paper, we used an accelerated failure time model with λ as defined above as the link function and T assumed to be Weibull. Let $\mu = -\ln \lambda = \boldsymbol{\beta}'\mathbf{X}$, $\alpha = 1/\sigma$, and $U = (\ln(T) - \mu)/\sigma$. Now μ is a location parameter and σ is a scale parameter. Then

$$f_U(u) = e^u e^{-e^u} \quad \text{and} \quad 1 - F_U(u) = e^{-e^u}. \tag{1}$$

If $\sigma > 1$, the hazard rate is decreasing; $\sigma < 1$ indicates that the hazard is increasing. (If $\sigma = 1$, the distribution is exponential.) In this paper, we assume that σ does not depend on $\mathbf{X}$; this need not be the case (see Anderson, 1991).

The usual likelihood function for right-censored data may be written

$$L_i = \prod_{i=1}^{n} \{f_i(t_i)^{\delta_{1i}}(1 - F_i(t_{1i}))^{1-\delta_{1i}}\}, \tag{2}$$

where

$$\delta_{1i} = \begin{cases} 1 & \text{if event at } t_i, \\ 0 & \text{if censored at } t_i, \end{cases}$$

$f_i(t_i)$ = density at time t_i given covariates $\mathbf{X}_i$,

$1 - F_i(t_{1i})$ = probability of survival beyond time t_{1i} (censoring point),

where survival is used to mean no event by time t_i. Here we have used f_i rather than the more explicit $f_{T|\mathbf{X}_i}$ as a shorthand notation.

To accommodate left- and interval censoring, we add two more components to the likelihood function, as follows. Let T_{0i} be a (possibly random) value such that if $T_{0i} > T_i$ then T_{0i} is observed and T_i is not. Similarly, let T_{1i} be a (possibly random) value such that if $T_{1i} < T_i$ then T_{1i} is observed and T_i is not. Let δ_{Ri} be an indicator that T_i is right-censored, $\{T_{0i} \leqslant T_i < T_{1i} = \infty\}$; let δ_{Li} be an indicator that T_i is left-censored, $\{0 = T_{0i} < T_i \leqslant T_{1i}\}$; and let δ_{Ii} be an indicator that T_i is interval-censored—that is, T_{0i} and T_{1i} are observed and T_i is not and $\{0 < T_{0i} < T_i \leqslant T_{1i} < \infty\}$. Finally, let δ_{Ei} be an indicator that T_i is observed exactly, i.e., $\delta_{Ei} = 1 - \delta_{Ri} - \delta_{Li} - \delta_{Ii}$. We assume that censoring is noninformative (Lawless, 1982). Now let t_{0i} be a realization of T_{0i} and let t_{1i} be a realization of T_{1i}. Then the portion of the likelihood that is a function of parameters can be written as

$$L = \prod_{i=1}^{n} \{f_i(t_i)^{\delta_{Ei}}F_i(t_{1i})^{\delta_{Li}}(1 - F_i(t_{0i}))^{\delta_{Ri}}(F_i(t_{1i}) - F_i(t_{0i}))^{\delta_{Ii}}\}. \tag{3}$$

Let $u_{ji} = (\ln t_{ji} - \mu_i)/\sigma$, for $j = 0, 1$, where $\mu_i = \boldsymbol{\beta}'\mathbf{X}_i$. Since $F_i(t) \equiv F_U((\ln t - \mu_i)/\sigma)$,

$$f_i(t) = \frac{d}{dt}F_i(t) = \frac{d}{dt}F_U\left(\frac{\ln t - \mu}{\sigma}\right) = f_U\left(\frac{\ln t - \mu}{\sigma}\right)\frac{1}{t\sigma}.$$

Then if $l = \ln L$,

$$l = \sum_{i=1}^{n} \{\delta_{Ei}(\ln f(u_i) - \ln \sigma - \ln t_i) + \delta_{Li}\ln(F(u_{1i})) + \delta_{Ri}\ln(1 - F(u_{0i})) + \delta_{Ii}\ln[F(u_{1i}) - F(u_{0i})]\}. \tag{4}$$

For the case of the Weibull distribution, from (1) and (4),

$$l = \sum_{i=1}^{n} \{\delta_{Ei}(u_i - e^{u_i} - \ln \sigma - \ln t_i) + \delta_{Li}\ln(1 - e^{u_{0i}}) - \delta_{Ri}e^{u_{1i}} + \delta_{Ii}\ln[e^{-e^{u_{0i}}} - e^{-e^{u_{1i}}}]\}.$$

Instead of estimating σ directly, we estimate γ, where $\sigma = e^{\gamma}$. Thus, the ranges for all parameters are $(-\infty, \infty)$. The MLEs are found using the Newton–Raphson procedure (see Kennedy and Gentle, 1980, pp. 442–449). This algorithm allows for changes in the usual step size and step direction to ensure that the likelihood increases with each iteration.

3. An Example

A usual procedure when event times are missing is to use likelihood function (2) with midpoints substituted for missing times. To see whether using likelihood function (4) gives significantly different results, a comparison was made using data on angina pectoris (AP) from the Framingham Offspring Study (see Feinleib et al., 1975). Estimates resulting from use of likelihood function (4) are hereafter called MLEs; estimates based on likelihood function (2) with midpoints substituted for interval-censored data will be called MDEs.

Data for the first three exams were available for study. Approximate time between the first and second exams was 8 years; between the second and third exams, 4 years. In both cases, actual dates of events were recorded if available; if not, the intervals of occurrence were noted, with times given for both endpoints. In most cases, these endpoints were exam dates.

Of the 2,568 females in the study free of AP at exam 1, 37 acquired AP before the third exam. Of these 37, 8 had event times recorded exactly, 16 others had the event in the first interval, and 13 others experienced it in the second. Of the total event times, 78.4% were missing.

Parameter estimates with one covariate, age, and their standard errors (s.e.) are displayed in Table 1. Note that $\gamma = -.5567$ implies $\sigma = .573$. For each parameter, the MDE is almost one-half of a standard deviation from the MLE. In absolute value, the MDE is greater than the MLE for the location and age parameters; for the scale parameter, γ, it is lower (and hence higher for σ). As expected, age is a highly significant predictor for AP. No other covariates were found to be significant, due no doubt to the relatively small number of events. Adding the natural logarithm of systolic blood pressure (SBP) to the model did not markedly change patterns in the original covariates. The difference between MLE and MDE estimates for the blood pressure parameter is in the expected direction.

Table 1
Angina pectoris: Females, one covariate

	MLE	s.e.	MDE	s.d.
Scale (γ)	−.5567	.215	−.4725	.1615
Location (β_0)	7.297	1.112	7.809	.9694
Age (β_1)	−.0582	.0159	−.0631	.0148

When the procedure was repeated for males, differences between estimates were small. Such differences as existed were opposite in sign from the case for females, and σ was higher (>.7). The percentage of missing times was much lower (<25%).

These results are typical, as will be seen in the simulation studies that follow.

4. Simulation

4.1 *Method*

In simulating the accelerated failure time model, we first used a single fixed interval (0, 1), where $t = 0$ was assumed to be the starting point and $t = 1$, the censoring point. A uniform(−15, 15) covariate $\mathbf{X}_1$, called "age," was included in the simulation, giving

$$\boldsymbol{\beta}'\mathbf{X} = \beta_0 + \beta_1 X_1.$$

Four values of σ were chosen, representing a range of hazard rates: $\sigma = .5$ (rapidly increasing), $\sigma = .8$ (increasing), $\sigma = 1.2$ (decreasing), and $\sigma = 2.0$ (rapidly decreasing). Probabilities of an event occurring for $X_1 = -15$ and $X_1 = 15$ were also selected. This yields two equations in two unknowns, which can be solved for values of β_0 and β_1. For example, if we assume that 30% of individuals of age 15 will get the event by time $T = 1$ but only 10% of those aged −15 will get it, we can write

$$\beta_0 + 15\beta_1 = -\sigma \log(-\log(.7)),$$

$$\beta_0 - 15\beta_1 = -\sigma \log(-\log(.9)),$$

and solve for β_0 and β_1 for given σ.

One hundred simulations were run for sample sizes 100, 200, 400, 600, 1,200, and 2,400, for occurrence rates 5–10 (5% for age −15, 10% for age 15), 10–30, 30–50, 30–70, and 80–90 for each of the four values of σ. The interval censoring was presumed to be noninformative and was determined by an independent (pseudo)-random Bernoulli variable. In most cases, it was assumed that 50% of event times were missing and 50% known exactly, though a few simulations with 25% and 75% missing were also tried. Each simulation was run in two ways, using the Newton–Raphson procedure for estimating parameters:

1. Equation (4) for the likelihood function was used, thus providing interval-based maximum likelihood estimates (MLEs).
2. Equation (2) was used, with midpoints substituted for interval-censored times. The resulting estimates are called MDEs.

No serious problems were encountered in running the simulations. In a very few instances, where percent-occurred was small, it was necessary to provide approximate starting values for the MLE estimates. These did not have to be at all exact; for example, using the MDE estimates was always successful. In every case, convergence was unique and quickly achieved.

For each simulation, average bias of the estimate was calculated. It was expected that variation of estimates would also be important since one would expect MDEs to have smaller variation than estimates from a sample with no interval censoring. To take variation as well as bias into account, mean squared errors (MSEs) were determined for each case.

4.2 *Results and Discussion*

We focus first on results for β_1, the age parameter. Tables 2 and 3 give examples. For complete results, see Odell (unpublished Ph.D. dissertation, Boston University, 1990).

Table 2 gives first an example for a sample size of 600 with a percent-occurred of 10–30, and then one for size 400, percent-occurred 80–90. Note that for rapidly increasing hazard ($\sigma = .5$), the bias of the MDE of β_1 is negative and, in absolute value, approximately equal to one-half of the MLE standard deviation. This result is similar for larger n, smaller percent-occurred, confirming results in the Framingham example of AP in women. Bias of

Table 2
Age parameter estimates and errors

Hazard	β_1		Est. β_1	Bias	s.d.	$\sqrt{\text{MSE}}$
		n = 600, % occurred = 10–30				
Rapidly decreasing	−.08130	MLE	−.07918	.00212	.02312	.02310
		MDE	−.05226	.02904	.01514	.03272
Decreasing	−.04878	MLE	−.04751	.00127	.01387	.01386
		MDE	−.03666	.01212	.01030	.01587
Increasing	−.03252	MLE	−.03167	.00085	.00925	.00924
		MDE	−.02888	.00364	.00796	.00871
Rapidly increasing	−.02032	MLE	−.01978	.00054	.00578	.00578
		MDE	−.02301	.00269	.00626	.00678
		n = 400, % occurred = 80–90				
Rapidly decreasing	−.02388	MLE	−.02116	.00272	.01267	.01290
		MDE	−.01021	.01367	.00659	.01516
Decreasing	−.01433	MLE	−.01270	.00163	.00760	.00774
		MDE	−.00737	.00696	.00451	.00828
Increasing	−.00955	MLE	−.00846	.00109	.00507	.00516
		MDE	−.00597	.00358	.00357	.00504
Rapidly increasing	−.00597	MLE	−.00529	.00070	.00317	.00323
		MDE	−.00482	.00115	.00288	.00308

% interval-censored = 50. Est. β_1, s.d., and MSE are the sample mean, standard deviation, and mean squared error, respectively, of simulation estimates of β_1, the age parameter. Rapidly decreasing: $\sigma = 2.0$; decreasing: $\sigma = 1.2$; increasing: $\sigma = .8$; rapidly increasing: $\sigma = .5$.

the MDE for more mildly increasing hazard ($\sigma = .8$) was much lower and positive, as was the case in the example of AP in men.

The examples in Table 2 indicate that the relative biases of MDEs are, in general, much more pronounced for larger values of β_1 (which correspond to larger values of σ). Example 1 also shows that for rapidly increasing hazard, the standard deviation of the MDE is higher than that of the MLE for small percent-occurred. This is contrary to expectation.

For smaller samples, relatively large MSEs tend to offset possible advantages of the MLE as compared to the MDE for estimating β_1. But as sample size increases, variation for the MLE decreases much more quickly than for the MDE. Biases of the MDE are fairly constant. Table 3 gives an example of the effects of sample size for decreasing hazard, percent-occurred = 30/70. As in this example, for $n \geq 400$, the MLE becomes superior to the MDE in terms of MSE in many cases. Table 4 summarizes where the MLE is superior for each n and each type of hazard function.

It is important to keep in mind the actual as well as comparative sizes of biases and MSEs. In general, both become larger as the hazard rate increases and, for given hazard, as the percent-occurred decreases.

These observations lead to another method for comparing MLEs and MDEs. For $\alpha = .05$, 95% of the confidence intervals $[\hat{\beta}_1 \pm 1.96\text{s.e.}(\hat{\beta}_1)]$ should include the true β_1. Here

$$\left.\frac{\partial^2 l}{\partial \beta_1^2}\right|_{\beta_1=\hat{\beta}_1}$$

is used as the estimate of the variance of $\hat{\beta}_1$. This "asymptotic" estimate, which is produced in the course of the Newton–Raphson procedure, is usually very close to the sample variance; any slight error tends to be in the direction of overestimation.

Table 3
Age parameter estimates, varying sample size

$\beta_1 = -.04866$		Est. β_1	Bias	s.d.	$\sqrt{\text{MSE}}$
$n = 100$	MLE	−.04756	.00110	.02122	.02114
	MDE	−.03459	.01407	.01487	.02041
$n = 200$	MLE	−.05249	.00383	.01501	.01542
	MDE	−.03790	.01076	.00994	.01461
$n = 400$	MLE	−.04821	.00045	.01153	.01148
	MDE	−.03499	.01367	.00783	.01573
$n = 600$	MLE	−.04836	.00030	.00987	.00983
	MDE	−.03497	.01369	.00677	.01524
$n = 1{,}200$	MLE	−.04858	.00008	.00652	.00649
	MDE	−.03505	.01361	.00468	.01438

Increasing hazard ($\sigma = 1.2$), percent-occurred = 30/70. Est. β_1, s.d., and MSE are the sample mean, standard deviation, and mean squared error, respectively, of simulation estimates of β_1.

Table 4
Circumstances where the MLE *is superior to the* MDE

n	80–90	30–70	30–50	10–30	5–10	80–90	30–70	30–50	10–30	5–10
			$\sigma = 2.0$					**$\sigma = 1.2$**		
100	*									
200	*	=					*			
400	*	*				*	*			
600	*	*	*	*		*	*			
1,200	*	*	*	*		*	*	*	*	
2,400	*	*	*	*	*	*	*	*	*	=
			$\sigma = .8$					**$\sigma = .5$**		
100								*		*
200							*	*	=	*
400						=	*	*	*	*
600		*				=	*	*	*	*
1,200	*	*	*			=	*	*	*	*
2,400	*	*	*	*		=	*	*	*	*

* indicates MLE superior to MDE in terms of MSE (mean squared error), where MLE = interval-based estimate, MDE = midpoint-based estimate. A blank space indicates MDE superiority; = indicates equivalence. % occurred is range of percent-occurred across age range. n is size of simulated sample. It should be kept in mind that the MLE is always superior to the MDE in terms of bias.

For MLEs, the requirement is satisfied virtually all the time. Results are shown in Table 5 for MDEs. In general, with decreasing failure rates ($\sigma > 1$), the confidence intervals are too short. For increasing failure rates ($\sigma < 1$), confidence intervals appear quite satisfactory unless a large number of events is expected.

Results for σ and β_0, the location parameter, were also obtained. (As mentioned earlier, the actual parameter estimated was $\gamma = \ln \sigma$, but for convenience of interpretation results are given in terms of σ.) For decreasing or mildly increasing hazard, MDEs overestimate σ and, in absolute value, underestimate β_0. For rapidly increasing hazard, errors are in the opposite directions for n fairly large or percent occurred relatively low.

This exactly parallels the Section 3 examples. For estimating σ and β_0, the MLE is superior to the MDE in terms of MSE for most hazard rates and percents-occurred, even for n as low as 100. For high percent-occurred and/or decreasing failure rate, the superiority is very marked.

Table 5
Percentages of inadequate confidence intervals for MDE

	% occurred	n = 400, % outside C.I.	n = 1,200, % outside C.I.
Rapidly decreasing	80–90	29	83
	10–30	29	66
	5–10	6	20
Decreasing	80–90	20	51
	10–30	12	38
	5–10	6	11
Increasing	80–90	4	27
	10–30	5	10
	5–10	4	8
Rapidly increasing	80–90	5	14
	10–30	5	11
	5–10	4	4

Average percent of simulated confidence intervals for MDE not covering true value of the age parameter, α = .05, based on 100 simulations. Rapidly decreasing: σ = 2.0; decreasing: σ = 1.2; increasing: σ = .8; rapidly increasing: σ = .5.

Simulations with only 25% of event times interval-censored gave results for MLEs and MDEs in directions similar to those for 50% missing. But, not surprisingly, MDEs in this case were closer to estimates based on a full sample (none missing). Hence, for 25% missing, the MLE is likely to be significantly better than the MDE only in those cases where its superiority for 50% missing is strongest. For 75% missing, conclusions for 50% missing are strengthened slightly.

Simulations with the interval divided into two smaller intervals were also run. All results were similar to those for the single-interval case, but the biases and standard deviations were somewhat smaller. Clearly, when intervals are very small, very little information is missing, and there is no need for MLEs.

4.3 *Conclusions*

When confronted with real interval-censored data, how does one decide whether it is worthwhile to use MLEs rather than MDEs when estimating parameters? Much depends on the purpose for which the estimates will be used and the resources available. Each situation requires its own appraisal of how much bias and mean squared error is tolerable. We can, however, make some general recommendations.

If estimates for the location and scale parameters are required, and resources permit, the MLE is almost always superior and worth using, especially for decreasing hazards. If interest lies only in the covariate parameter(s), one must first consider the percentage of missing data, the size of the sample, and the width of the intervals. If all are relatively small, there is no reason not to use the MDE. Even if the intervals are relatively wide, the high variance of the MLE for small samples will offset its superiority in terms of average bias. Similarly, if less than 25% of the data is missing, the MDE will usually be adequate even for fairly large data sets.

If the sample is large ($n > 400$), then one must have some idea of the hazard rate in order to make a decision. This can be loosely estimated by graphing the data or by making a trial run using MDEs. An MDE for the scale parameter of .85 or greater corresponds to a moderately decreasing failure rate ($\sigma > 1.2$); an MDE estimate of 1.25 or more indicates a rapidly decreasing hazard ($\sigma > 2.0$). An MDE between .6 and .8 implies a moderately increasing to flat hazard; MDEs below .6 correspond to a rapidly decreasing hazard ($\sigma < .5$).

If the hazard rate appears flat or only slowly increasing or decreasing, then the MDE is adequate. The steeper the hazard, the greater the need for the MLE. This is particularly true for very large samples.

ACKNOWLEDGEMENTS

The work of the first author was supported by National Heart, Lung, and Blood Institute Grant R01HL41423-02.

RÉSUMÉ

Le modèle de regression pour temps de survie accéléré est utilisé en général pour des données censurées à droite. Cet article étudie l'utilisation d'un modèle de régression pour temps de survie accéléré basé sur la distribution de Weibull lorsque les données peuvent être censurées par intervalles ou censurées à gauche. Deux méthodes d'analyse sont considérées. En premier lieu, les estimateurs du maximum de vraisemblance (MLE) sont calculés pour la répartition observées des données censurées. Ces estimateurs sont comparés avec les estimateurs obtenus en remplaçant les données censurées par intervalle ou à gauche par des censures en milieu d'intervalle (MDE). Des simulations indiquent que pour des échantillons relativement grands, il arrive souvent que le MLE soit mieux que le MDE. Pour des échantillons dans lesquels la fonction de risque est à peu près constante, ou dans lesquels le pourcentage de données censurées par intervalle est petit, le MDE est adéquat. On discute un exemple qui utilise les données de l'Étude de Framingham.

REFERENCES

Anderson, K. M. (1991). A nonproportional hazards Weibull accelerated failure time regression model. *Biometrics* **47,** 281–288.

Brookmeyer, R. and Goedert, J. J. (1989). Censoring in an epidemic with an application to hemophilia-associated AIDS. *Biometrics* **45,** 325–335.

Feinleib, M., Kannel, W. B., Garrison, R. J., et al. (1975). The Framingham Offspring Study: Design and preliminary data. *Preventive Medicine* **4,** 518.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42,** 845–854.

Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41,** 933–945.

Flygare, M. E., Austin, J. A., and Buckwater, R. M. (1985). Maximum likelihood estimation for the two-parameter Weibull distribution based on interval data. *IEEE Transactions on Reliability* **34,** 57–59.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kennedy, J. E. and Gentle, W. J. (1980). *Statistical Computing.* New York: Dekker.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Self, S. G. and Grossman, E. A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* **42,** 521–530.

SAS/STAT Users Guide, Version 6, 4*th edition* (1990). Cary, North Carolina: SAS Institute Inc.

Received December 1989; *revised November* 1990, *June* 1991, *and January* 1992; *accepted February* 1992.