Biometrika (1996), **83**, 4, *pp.* 747–761
Printed in Great Britain

Estimation from truncated lifetime data with supplementary information on covariates and censoring times

BY X. JOAN HU AND JERALD F. LAWLESS
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

SUMMARY

In epidemiology, reliability and other areas one encounters situations where response times and covariates for individuals in a population are observed only if the response time does not exceed an associated censoring time. Methods for obtaining supplementary information about covariates and censoring times for nonresponding individuals are considered. Estimation procedures are developed for the resulting combination of truncated data and supplementary data. Theoretical and simulation results are given, and the methodology illustrated on some automobile failure data collected under a warranty plan.

Some key words: Epidemiology; Estimating function; Field reliability; Missing data; Pseudo likelihood.

1. INTRODUCTION

Response-biased data are common in epidemiology, field reliability and other areas. A familiar example is when an individual from some finite population $\mathscr{P}$ is observed, and a response variable T_i and covariate Z_i measured, if and only if T_i satisfies some condition. Although the ideas presented apply more generally, we will focus here on the situation where T_i represents a lifetime or response time of some sort. There is also a truncation or censoring time $\mathscr{T}_i$ associated with each individual, and individual i in $\mathscr{P}$ is observed, and T_i, $\mathscr{T}_i$, Z_i are obtained, if $T_i \leqslant \mathscr{T}_i$. For individuals with $T_i > \mathscr{T}_i$ none of T_i, $\mathscr{T}_i$ or Z_i are available, but we consider the possibility of obtaining supplementary information on $(\mathscr{T}_i, Z_i)$ values for such individuals, as described later. Our interest is in estimating the distribution of T given $Z = z$ or, slightly more generally, the distribution $F(t \mid \tau, z)$ of T given $Z = z$ and $\mathscr{T} = \tau$ for the process which generates the population data.

A comment on terminology may be helpful. If no supplementary information is available and we know only about the individuals with $T_i \leqslant \mathscr{T}_i$ then the data are usually called truncated and the $\mathscr{T}_i$'s are referred to as truncation times, e.g. Woodroofe (1985), Kalbfleisch & Lawless (1989). If data on individuals with $T_i > \mathscr{T}_i$ are available, however, we usually refer to the $\mathscr{T}_i$'s as censoring times. In this paper the number of individuals with $T_i > \mathscr{T}_i$ is assumed known, but not necessarily the $\mathscr{T}_i$ or Z_i values. Nevertheless, it seems more accurate to refer to the $\mathscr{T}_i$'s as censoring times, and we shall do this here.

Problems falling into this framework arise in epidemiology and field reliability. For example, for cases of childhood diabetes in a population the age at onset T_i and covariates Z_i may be routinely obtained. For any individual in the population there will be a censoring time, i.e. age, τ_i: if the individual is born into the population at calendar time d_i and the current calendar time is D, then $\tau_i = \min(D - d_i, w_i)$, where w_i represents the age at withdrawal of the individual from the population due, for example, to migration, death, or leaving the age group under study. Individuals who migrate into the population after birth also create a truncation time on the left; the methods in the paper may be extended to deal with this but we will ignore it in what follows. Similar problems arise with the estimation of a failure time distribution from warranty or field return data on manufactured items. Suppose an item is sold at calendar time d_i and covered by a warranty or field return scheme which ensures that a failure occurring before time $d_i + W$ is reported. If we consider the population of items sold by time D, then the time T_i to failure for item i is observed, and covariates Z_i measured, if and only if $T_i \leqslant \tau_i$, where $\tau_i = \min(D - d_i, W)$.

If the only information available is the observed values $(T_i, \mathscr{T}_i, Z_i) = (t_i, \tau_i, z_i)$ for individuals in the population who satisfy $t_i \leqslant \tau_i$, inferences about the parameters θ in a parametric specification $F(t \mid \theta; \tau, z)$ may be uninformative, e.g. Kalbfleisch & Lawless (1988a, 1989). However, if the population size M is known at least approximately and if supplementary information about the distribution of $\mathscr{T}$ and Z in $\mathscr{P}$ is available, then much more precise inference about θ may be possible. This has been explored in a number of specific contexts. Suzuki (1985a, b, 1987, 1993, 1995), and K. Suzuki and T. Kasashima, in an unpublished University of Electro-Communications Department of Communications and Systems Engineering, Japan report 'Estimation of lifetime distribution from incomplete field data with different observational periods', have considered supplementary random sampling of unfailed items in connection with industrial field failures when there are no covariates, and Kalbfleisch & Lawless (1988a, b) presented extensions to deal with covariates. Related problems have been considered for categorical responses by Hsieh, Manski & McFadden (1985), Wild (1991) and others who use information on population stratum sizes to supplement case-control or choice-based samples, and by Jewell (1985) and others in connection with response-biased sampling in regression. Prentice's case-cohort design (1986) in epidemiology, whereby response time and covariate information on 'cases', i.e. all individuals with failure or response times t_i satisfying $t_i \leqslant \tau_i$, are supplemented by information on all individuals in a randomly selected cohort, is similar.

The current paper deals with a general framework where the distribution $F(t \mid \theta; \tau, z)$ of T given $\mathscr{T} = \tau$ and $Z = z$ is specified parametrically, and our interest is in estimating θ, but the distribution of $\mathscr{T}$ and Z is treated nonparametrically. We consider two important types of supplementary information about $\mathscr{T}$ and Z: (i) follow-up sampling of unfailed, i.e. nonresponding, individuals; (ii) separate survey information. Pseudo likelihood methods which extend those of Kalbfleisch & Lawless (1988a, b) are presented for estimating θ, and we demonstrate the gains in efficiency achieved by using supplementary information. The paper is organised as follows. In § 2 we present pseudo likelihood estimating functions based on truncated data and supplementary information. Implementations of the methods are given for the two types of supplementary information mentioned above. Section 3 gives asymptotic properties of the proposed estimators, and estimates of their asymptotic variances. Section 4 presents simulation results that show finite sample properties and address the adequacy of asymptotic approximations from § 3. An application of the methodology to some automobile warranty data is discussed in § 5. Section 6 gives concluding remarks.

2. Estimating functions with supplementary data

Consider a population $\mathscr{P}$ of known size M, and suppose that (t_i, τ_i, z_i) $(i = 1, \ldots, M)$ arise as a random sample from a distribution with joint probability density function

$$f(t \mid \theta; \tau, z)\, dG(\tau, z) \quad (t > 0, (\tau, z) \in \mathscr{A}). \tag{1}$$

Usually T and $\mathscr{T}$ are independent given Z but sometimes, for example when there is a trend in response times over calendar time in the examples of § 1, we may wish to allow for dependence. Consequently we retain this in our notation for the density $f(t|\theta;\tau,z)$ of T given the censoring time τ and a q-vector of covariates z, which we assume is specified up to a p-vector of parameters θ. The function $G(\tau,z)$ is an arbitrary cumulative distribution function of $(\mathscr{T},Z)$ with range $\mathscr{A}\subseteq\mathscr{R}^{q+1}$. Our main interest is in estimation of θ, while treating $G(\tau,z)$ nonparametrically, since a model for $G(\tau,z)$ is often hard to specify. Denote the cumulative distribution function of T given (τ,z) by

$$F(t|\theta;\tau,z)=\mathrm{pr}(T\leqslant t|\tau,z;\theta),$$

and the survivor function by

$$\bar{F}(t|\theta;\tau,z)=1-F(t|\theta;\tau,z).$$

By convention, all vectors are column-vectors and if $\theta=(\theta_1,\ldots,\theta_p)'$ and $g(\theta)$ is a r-vector then $\partial g(\theta)/\partial\theta$ is a $r\times p$ matrix with (i,j) entry $\partial g_i(\theta)/\partial\theta_j$.

The ith individual is 'observed', and t_i, τ_i and z_i are determined, if and only if $t_i\leqslant\tau_i$. For the other individuals we know only that $t_i>\tau_i$. The primary data are thus $\mathcal{O}_1\cup\mathcal{O}_2$, where

$$\mathcal{O}_1=\{(t_i,\tau_i,z_i):i\in\mathscr{P}^*\},\quad \mathcal{O}_2=\{i:i\in\mathscr{P};\ t_i>\tau_i\},$$

with $\mathscr{P}^*=\{i:i\in\mathscr{P};\ t_i\leqslant\tau_i\}$ of size m.

If only $\mathcal{O}_1$ is known, inferences about θ can be based on the truncated conditional likelihood function

$$L_T(\theta)=\prod_{i\in\mathscr{P}^*}\frac{f(t_i|\theta;\tau_i,z_i)}{F(\tau_i|\theta;\tau_i,z_i)}. \tag{2}$$

As illustrated in Kalbfleisch & Lawless (1988a) and Hu & Lawless (1996), the likelihood function (2) can be uninformative about θ when m/M is small. Information about the number of unobserved items $M-m$ and their (τ,z) values is in that case very valuable, and we now consider such information.

We now consider an approach which provides estimating functions for θ based on $\mathcal{O}_1$, $\mathcal{O}_2$ and different types of supplementary data. If the values of (τ_i,z_i) were known for the individuals in $\mathcal{O}_2$, we could use the familiar censored data likelihood

$$\begin{aligned} L_C(\theta)&:=\prod_{i\in\mathscr{P}^*}f(t_i|\theta;\tau_i,z_i)\,dG(\tau_i,z_i)\prod_{i\in\mathcal{O}_2}\bar{F}(\tau_i|\theta;\tau_i,z_i)\,dG(\tau_i,z_i)\\ &\propto\prod_{i\in\mathscr{P}^*}f(t_i|\theta;\tau_i,z_i)\prod_{i\in\mathcal{O}_2}\bar{F}(\tau_i|\theta;\tau_i,z_i). \end{aligned}\tag{3}$$

This yields the score vector $S_C(\theta):=\partial\log L_C(\theta)/\partial\theta$, which we write as

$$S_C(\theta)=\sum_{i\in\mathscr{P}^*}\left\{\frac{\partial\log f(t_i|\theta;\tau_i,z_i)}{\partial\theta}-\frac{\partial\log\bar{F}(\tau_i|\theta;\tau_i,z_i)}{\partial\theta}\right\}+W(\theta),$$

where $W(\theta)=\sum_{i\in\mathscr{P}}\partial\log\bar{F}(\tau_i|\theta;\tau_i,z_i)/\partial\theta$. Our approach is to use supplementary information, described below, to estimate $W(\theta)$ by $\tilde{W}(\theta)$, and then to use the estimating function

$$PS\{\theta;\tilde{W}(\theta)\}:=\sum_{i\in\mathscr{P}^*}\left\{\frac{\partial\log f(t_i|\theta;\tau_i,z_i)}{\partial\theta}-\frac{\partial\log\bar{F}(\tau_i|\theta;\tau_i,z_i)}{\partial\theta}\right\}+\tilde{W}(\theta). \tag{4}$$

This is equivalent to estimating the component of $\log L_C(\theta)$ that has missing values of τ_i

and z_i. We may then estimate θ by solving $PS\{\theta;\ \tilde{W}(\theta)\}=0$; implementation and variance estimates are described in § 3. The function (4) may be termed a pseudo score or an estimated score function in the terminology of Kalbfleisch & Lawless (1988a, b). The idea has been used previously in various contexts, e.g. Suzuki (1985a), Kalbfleisch & Lawless (1988a, b), Pepe & Fleming (1991), Pepe (1992) and Reilly & Pepe (1995).

We consider two approaches to obtaining supplementary data, and thus estimating $W(\theta)$. One approach (Suzuki, 1985a, b; Kalbfleisch & Lawless, 1988a) is to sample randomly some individuals in $\mathcal{O}_2$, and observe their τ_i and z_i values. Let $\mathcal{S}^f$ denote the supplementary sample. We consider here a simple random sample without replacement of $n^f=p^f(M-m)$ items from the $M-m$ ones with $t_i>\tau_i$; p^f is a prespecified constant. This is the most frequently-used scheme, but the methods below may also be adapted to other schemes.

The term $W(\theta)$ is the population total in $\mathcal{P}$ for the variate $\partial \log \bar{F}(\tau\,|\,\theta;\ \tau, z)/\partial\theta$. Thus, $\tilde{W}(\theta)$ is taken here as

$$\tilde{W}_f(\theta):=\sum_{i\in\mathcal{P}^*}\frac{\partial \log \bar{F}(\tau_i\,|\,\theta;\ \tau_i, z_i)}{\partial\theta}+\frac{M-m}{n^f}\sum_{i\in\mathcal{S}^f}\frac{\partial \log \bar{F}(\tau_i\,|\,\theta;\ \tau_i, z_i)}{\partial\theta}. \tag{5}$$

The resulting estimating function (4) is a generalisation of the derivative of the pseudo-log-likelihood (3.2) in Kalbfleisch & Lawless (1988a). By defining $R_{1i}=I(i\in\mathcal{P}^*)$ and $R_{2i}=I(i\in\mathcal{S}^f)$ for $i=1,\ldots,M$ and rewriting (4) as

$$PS\{\theta;\ \tilde{W}_f(\theta)\}=\sum_{i\in\mathcal{P}}\left\{R_{1i}\frac{\partial \log f(t_i\,|\,\theta;\ \tau_i, z_i)}{\partial\theta}+\frac{R_{2i}}{p^f}\frac{\partial \log \bar{F}(\tau_i\,|\,\theta;\ \tau_i, z_i)}{\partial\theta}\right\}, \tag{6}$$

it is seen that $PS\{\theta;\ \tilde{W}_f(\theta)\}$ has expectation zero under model (1) and the supplementary sampling scheme. As outlined in § 3, it provides consistent estimation of θ.

The second approach is to obtain information about the distribution of (τ, z) from a random sample that may or may not be independent of the failure data. Suppose that (τ_j, z_j) for $j\in\mathcal{S}^s$ is a random sample from $G(\tau, z)$ with size $n^s=p^sM$. Often $\mathcal{S}^s$ will be selected from $\mathcal{P}$ but, if there is a process which generates the (τ_i, z_i)'s in $\mathcal{P}$ and other similar data, it may be possible to obtain a supplementary sample that is independent of $\mathcal{P}$. In either case we take $\tilde{W}(\theta)$ in (4) as

$$\tilde{W}_s(\theta):=\frac{M}{n^s}\sum_{i\in\mathcal{S}^s}\frac{\partial \log \bar{F}(\tau_j\,|\,\theta;\ \tau_j, z_j)}{\partial\theta}. \tag{7}$$

Writing (4) in the case as

$$PS\{\theta;\ \tilde{W}_s(\theta)\}\equiv S_C(\theta)-W(\theta)+\tilde{W}_s(\theta), \tag{8}$$

we see that $PS\{\theta;\ \tilde{W}_s(\theta)\}$ has expectation zero under (1) and the supplementary sampling scheme. As shown in § 3, it provides consistent estimation of θ.

Estimates $\hat{\theta}$ are obtained by solving $PS(\theta;\ \tilde{W}(\theta)\}=0$. We have used Newton's method and have not encountered any convergence difficulties. In § 3 we state asymptotic properties of $\hat{\theta}$ and provide variance estimates.

We conclude this section with two additional remarks. First, there are sometimes administrative records, perhaps in aggregate form, which give (τ_i, z_i) values in a population. For example, with human populations there may be records of birth dates and sex of individuals; for manufactured products under warranty there may be aggregate, e.g. monthly, sales data. In these situations a supplementary sample may be drawn directly from the records. Secondly, the population size M may be known only approximately, even though one may be able to select random samples from it. Since M figures in (5) and (7), it is advisable to check the sensitivity of estimation results to the value assumed for M.

3. Asymptotic results and variance estimates

In this section we state asymptotic results about estimates $\hat{\theta}$ obtained by setting (4) equal to zero. Then we consider the two types of supplementary sampling described in § 2, which lead to the use of (5) and (7), respectively, for $\tilde{W}(\theta)$ in (4). Asymptotic variance estimates for $\hat{\theta}$ are given for each case. Derivations of results are outlined in Appendix 1.

We define $\Delta(\theta)=\tilde{W}(\theta)-W(\theta)$, and let θ_0 be the true value of θ.

Theorem 1. *Under conditions stated in Appendix* 1, *for* $M\to\infty$, $\hat{\theta}$ *is a strongly consistent estimator of* θ_0, *and* $\sqrt{M}(\hat{\theta}-\theta_0)$ *converges in distribution to a multivariate normal random variable with mean zero and covariance matrix*

$$\Sigma^{-1}(\theta_0)+\Sigma^{-1}(\theta_0)\{2H(\theta_0)+K(\theta_0)\}\Sigma^{-1}(\theta_0), \tag{9}$$

where

$$\Sigma(\theta_0)=\lim_{M\to\infty}\operatorname{var}\{S_C(\theta_0)\}/M,\quad K(\theta_0)=\lim_{M\to\infty}\operatorname{var}\{\Delta(\theta_0)\}/M,$$

$$H(\theta_0)=\lim_{M\to\infty}\operatorname{cov}\{S_C(\theta_0),\Delta(\theta_0)\}/M.$$

It is shown in Appendix 2 that, with $\tilde{W}_f(\theta)$ and $\tilde{W}_s(\theta)$ based on (5) and (7), the resulting estimator $\hat{\theta}$ satisfies the conditions for Theorem 1. Estimates for the three matrices in (9) are given below for each case.

Supplementary follow-up data. Define

$$v_i(\theta)=\partial\log\bar{F}(\tau_i|\theta;\tau_i,z_i)/\partial\theta,\quad \bar{v}_f(\theta)=\sum_{i\in\mathscr{S}^f}v_i(\theta)/n^f.$$

It is shown in Appendix 2 that $H(\theta_0)=0$ and that consistent estimates of $\Sigma(\theta_0)$ and $K(\theta_0)$ are given by

$$\hat{\Sigma}_f:=-\frac{1}{M}\frac{\partial PS\{\theta;\tilde{W}_f(\theta)\}}{\partial\theta}\bigg|_{\theta=\hat{\theta}}, \tag{10}$$

$$\hat{K}_f:=\left(1-\frac{m}{M}\right)\left(\frac{1}{p^f}-1\right)\frac{1}{n^f-1}\sum_{i\in\mathscr{S}^f}\{v_i(\hat{\theta})-\bar{v}_f(\hat{\theta})\}\{v_i(\hat{\theta})-\bar{v}_f(\hat{\theta})\}'. \tag{11}$$

Supplementary random sample from $\mathscr{P}$. Let $\bar{v}_s(\theta)=\sum v_i(\theta)/n^s$, where the sum is over $i\in\mathscr{S}^s$. In this case $H(\theta_0)=0$, (10) with $\tilde{W}_s(\theta)$ replacing $\tilde{W}_f(\theta)$ consistently estimates $\Sigma(\theta_0)$, and $K(\theta_0)$ is estimated by

$$\hat{K}_{s1}:=\left(\frac{1}{p^f}-1\right)\frac{1}{n^s-1}\sum_{i\in\mathscr{S}^s}\{v_i(\hat{\theta})-\bar{v}_s(\hat{\theta})\}\{v_i(\hat{\theta})-\bar{v}_s(\hat{\theta})\}'. \tag{12}$$

Supplementary random sample independent of $\mathscr{P}$. Define

$$u_i(\theta)=\partial\log\{f(t_i|\theta;\tau_i,z_i)/\bar{F}(\tau_i|\theta;\tau_i,z_i)\}/\partial\theta.$$

Then (10) with $\tilde{W}_s(\theta)$ replacing $\tilde{W}_f(\theta)$ consistently estimates $\Sigma(\theta_0)$, and $H(\theta_0)$ and $K(\theta_0)$

are estimated by

$$\hat{K}_{s2} := \left(\frac{1}{p^f} + 1\right) \frac{1}{n^s - 1} \sum_{i \in \mathscr{S}^s} \{v_i(\hat{\theta}) - \bar{v}_s(\hat{\theta})\} \{v_i(\hat{\theta}) - \bar{v}_s(\hat{\theta})\}', \tag{13}$$

$$\hat{H}_{s2} := \frac{1}{M} \left\{ \sum_{i \in \mathscr{P}^*} u_i(\hat{\theta}) v_i(\hat{\theta})' - \sum_{i \in \mathscr{P}^*} u_i(\hat{\theta}) \bar{v}_s(\hat{\theta})' \right\} \tag{14}$$

$$+ \frac{1}{n^s - 1} \sum_{i \in \mathscr{S}^s} \{v_i(\hat{\theta}) - \bar{v}_s(\hat{\theta})\} \{v_i(\hat{\theta}) - \bar{v}_s(\hat{\theta})\}'. \tag{15}$$

Note that $\Sigma(\theta_0)^{-1}$ is the asymptotic covariance matrix for $\hat{\theta}_C$, the maximum likelihood estimator that is obtained by maximising (3), i.e. by solving the likelihood equation $S_C(\theta) = 0$, in the situation where the (τ_i, z_i) values are known for all individuals in $\mathscr{P}$. Thus the second term in (9) represents the loss of information entailed by having only samples of the (τ_i, z_i) values for individuals not in $\mathscr{P}^*$. Simulations in § 4 show that, in cases where truncation is heavy, our procedures with supplementary samples having p^f or p^s as small as 0·05 recover much of the missing information. An application is given in § 5.

4. SIMULATIONS

To study the behaviour of the proposed methods, we performed some simulations based on the Weibull proportional hazards model, a widely-used parametric lifetime regression model. We took censoring times and lifetimes to be independent and varied their distributions to give heavy to moderate truncation ($m/M = 0{\cdot}05$ to $0{\cdot}25$). For the present illustration, parameter values and censoring were chosen to be realistic for warranty and field reliability applications like one discussed in § 5.

We chose $M = 4000$, and generated lifetimes t_i $(i = 1, \ldots, M)$ from the probability density function

$$f(t \mid \tau_i, z_i) = \delta t^{\delta - 1} e^{\beta_0 + \beta_1 z_i} \exp(-t^\delta e^{\beta_0 + \beta_1 x_i}), \tag{16}$$

where the z_i's were independent and took on values 0 and 1 each with probability 0·5. A random sample y_i $(i = 1, \ldots, M)$ from $N(\mathscr{T}^0, \sigma^2)$ with $\mathscr{T}^0 = 300{\cdot}0$ and $\sigma = 80{\cdot}0$ was then drawn, and truncation times τ_i obtained by $\tau_i = \min(y_i, \mathscr{T}^0)$. The set $\{(t_i, \tau_i, z_i) : t_i \leqslant \tau_i\}$ is taken as the observed, truncated, data. The size of this set is denoted by m. In addition, supplementary data of the three types were simulated as follows.

Case 1. We randomly selected items with probability p^f from the $M - m$ items with $t_i > \tau_i$. The (τ_i, z_i)'s for the selected items play the role of a follow-up sample. We considered $p^f = 0{\cdot}05$, $0{\cdot}10$ and $0{\cdot}20$.

Case 2. A survey sample was chosen by selecting n^s pairs of (τ_j^*, z_j^*) randomly from $\{(\tau_i, z_i),\ i = 1, \ldots, M\}$. We determined n^s as Mp^s with p^s being 0·05, 0·10 and 0·20.

Case 3. Values τ_j^* and z_j^* $(j = 1, \ldots, n^s)$ were generated in the same way as $\{(\tau_i, z_i),\ i = 1, \ldots, M\}$. The set $\{(\tau_j^*, z_j^*) : j = 1, \ldots, n^s\}$ was taken as a survey sample independent of the M items.

We investigated various situations: in each of them, the observed data were supplemented by one of the three types of additional information described above. For each of the cases, the estimating function (4) was used to estimate the parameters, and to compute estimates of the asymptotic covariance matrices for $\surd M(\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1, \hat{\delta} - \delta)$, as discussed in § 3. In our simulations we also investigated maximum likelihood estimation based on the likelihoods L_T, from (2), and L_C, from (3). The likelihood L_T, based solely on the truncated data, is relatively uninformative about the parameters, in many cases giving standard deviations an order of magnitude larger than the other methods, which utilise supplementary information. This agrees with results of Kalbfleisch & Lawless (1988a) in a simpler setting. Since L_C represents the 'full information' situation, comparison with it indicates how far (4) with supplementary data goes toward restoring the missing information due to truncation. To conserve space we do not present the simulation results for L_T but give those for L_C, and we show results only for p^f, $p^s = 0{\cdot}05$ and $0{\cdot}20$.

We give results for three models, of the form (16) with $(\beta_0, \beta_1, \delta) = (-8{\cdot}6, 1{\cdot}0, 1{\cdot}0)$, $(-15{\cdot}8, 1{\cdot}0, 2{\cdot}5)$, $(-17{\cdot}6, 1{\cdot}0, 2{\cdot}5)$. These give approximate proportions $m/M = 0{\cdot}10$, $0{\cdot}25$, $0{\cdot}05$, respectively. In the tables we use PS_f, PS_{s1} and PS_{s2} to indicate the results obtained by applying the methods based on (4) with a follow-up sample, a random sample from $\mathscr{P}$, and an independent random sample, respectively. The results shown in the tables are all based on 100 repetitions of the simulation.

The sample means of the parameter estimates obtained by different methods are shown in Table 1(a). Table 1(b) gives the sample standard deviations of the estimates. All of the

Table 1. *Sample means and sample standard deviations of the estimates under three cases*

(*a*) *Sample means*

Methods		Case 1 β_0	Case 1 β_1	Case 1 δ	Case 2 β_0	Case 2 β_1	Case 2 δ	Case 3 β_0	Case 3 β_1	Case 3 δ
L_C		−8·641	1·000	1·006	−15·769	0·997	2·496	−17·685	0·989	2·518
PS_f	$(p^f = 0{\cdot}20)$	−8·641	0·997	1·007	−15·768	0·990	2·497	−17·718	0·984	2·523
	$(p^f = 0{\cdot}05)$	−8·634	1·003	1·005	−15·794	0·991	2·501	−17·749	0·989	2·529
PS_{s1}	$(p^s = 0{\cdot}20)$	−8·634	1·005	1·007	−15·779	0·997	2·500	−17·674	0·979	2·517
	$(p^s = 0{\cdot}05)$	−8·668	1·018	1·007	−15·825	1·009	2·503	−17·692	0·986	2·515
PS_{s2}	$(p^s = 0{\cdot}20)$	−8·635	0·986	1·007	−15·782	1·005	2·497	−17·720	1·000	2·521
	$(p^s = 0{\cdot}05)$	−8·648	1·000	1·008	−15·808	1·029	2·500	−17·730	0·986	2·524

(*b*) *Sample standard deviations*

Methods		Case 1 $\hat{\sigma}_0$	Case 1 $\hat{\sigma}_1$	Case 1 $\hat{\sigma}_2$	Case 2 $\hat{\sigma}_0$	Case 2 $\hat{\sigma}_1$	Case 2 $\hat{\sigma}_2$	Case 3 $\hat{\sigma}_0$	Case 3 $\hat{\sigma}_1$	Case 3 $\hat{\sigma}_2$
L_C		0·334	0·107	0·058	0·363	0·063	0·065	0·875	0·141	0·146
PS_f	$(p^f = 0{\cdot}20)$	0·334	0·122	0·057	0·402	0·102	0·071	0·875	0·154	0·150
	$(p^f = 0{\cdot}05)$	0·342	0·180	0·057	0·492	0·159	0·087	0·963	0·202	0·165
PS_{s1}	$(p^s = 0{\cdot}20)$	0·339	0·124	0·059	0·394	0·107	0·069	0·881	0·167	0·150
	$(p^s = 0{\cdot}05)$	0·350	0·181	0·058	0·497	0·181	0·089	0·916	0·195	0·154
PS_{s2}	$(p^s = 0{\cdot}20)$	0·345	0·130	0·058	0·402	0·117	0·071	0·850	0·173	0·146
	$(p^s = 0{\cdot}05)$	0·260	0·193	0·059	0·499	0·200	0·084	0·856	0·211	0·145

Case 1, $(\beta_0, \beta_1, \delta) = (-8{\cdot}6, 1{\cdot}0, 1{\cdot}0)$; Case 2, $(\beta_0, \beta_1, \delta) = (-15{\cdot}8, 1{\cdot}0, 2{\cdot}5)$; Case 3, $(\beta_0, \beta_1, \delta) = (-17{\cdot}6, 1{\cdot}0, 2{\cdot}5)$.

$\hat{\sigma}_0 = \{\bar{\mathrm{var}}(\hat{\beta}_0)\}^{1/2}$, $\hat{\sigma}_1 = \{\bar{\mathrm{var}}(\hat{\beta}_1)\}^{1/2}$, $\hat{\sigma}_2 = \{\bar{\mathrm{var}}(\hat{\delta})\}^{1/2}$, where vār is the sample variance from the 100 repetitions.

estimators have small bias relative to the standard deviations. Table 1(b) shows that using even a 5% supplementary sample, p^f or $p^s = 0{\cdot}05$, gives informative inferences. Once the supplementary sample fraction rises to 20%, the estimates for β_0 and δ based on (4) are very nearly as efficient as those based on L_C. The estimates of the regression coefficient β_1 have standard deviations about 25% higher than those from L_C when m/M is smaller, for Cases 1 and 3, and about 80–90% higher when m/M is larger, for Case 2. Table 1(b) also shows that increasing the size of the supplementary sample makes a bigger difference for estimation of the regression coefficient β_1 than for the parameters β_0 and δ in (16). These results are in general agreement with simulation results of Kalbfleisch & Lawless (1988a).

Table 2. *Comparison of sample variances and variance estimates*

*Methods		Case 1			Case 2			Case 3		
		β_0	β_1	δ	β_0	β_1	δ	β_0	β_1	δ
L_C	$\bar{\text{v}}\text{ar}$	0·1152	0·0115	0·0034	0·1320	0·0040	0·0042	0·7655	0·0200	0·0214
	$\text{m}\hat{\text{v}}\text{ar}$	0·0979	0·0145	0·0028	0·1617	0·0046	0·0050	0·9174	0·0251	0·0283
$PS_f\ (p^f = 0{\cdot}10)$	$\bar{\text{v}}\text{ar}$	0·1138	0·0209	0·0033	0·1880	0·0175	0·0058	0·8455	0·0282	0·0245
	$\text{m}\hat{\text{v}}\text{ar}$	0·1022	0·0239	0·0028	0·1990	0·0145	0·0061	0·9495	0·0354	0·0293
$PS_{s1}\ (p^s = 0{\cdot}10)$	$\bar{\text{v}}\text{ar}$	0·1178	0·0200	0·0034	0·1839	0·0204	0·0057	0·7798	0·0314	0·0220
	$\text{m}\hat{\text{v}}\text{ar}$	0·1027	0·0247	0·0029	0·2182	0·0190	0·0066	0·9461	0·0357	0·0292
$PS_{s2}\ (p^s = 0{\cdot}10)$	$\bar{\text{v}}\text{ar}$	0·1224	0·0231	0·0034	0·1967	0·0232	0·0055	0·7952	0·0324	0·0243
	$\text{m}\hat{\text{v}}\text{ar}$	0·1038	0·0270	0·0029	0·2341	0·0224	0·0070	0·9565	0·0381	0·0295

Case 1, $(\beta_0, \beta_1, \delta) = (-8{\cdot}6, 1{\cdot}0, 1{\cdot}0)$; Case 2, $(\beta_0, \beta_1, \delta) = (-15{\cdot}8, 1{\cdot}0, 2{\cdot}5)$; Case 3, $(\beta_0, \beta_1, \delta) = (-17{\cdot}6, 1{\cdot}0, 2{\cdot}5)$.
$\bar{\text{v}}\text{ar}$, sample variance from the 100 repetitions; $\text{m}\hat{\text{v}}\text{ar}$, sample mean of the variance estimates from the 100 repetitions.

Tables 2 and 3 address the adequacy of variance estimation and the asymptotic normal approximations of § 3, which one would use to get confidence intervals or tests for parameters. We give in Table 2 the sample variances of the parameter estimates, and the sample means of the variance estimates from the 100 repetitions obtained for each of the cases on which Table 1 is based. To conserve space only results for p^f and $p^s = 0{\cdot}10$ are given. The sample variances are relatively close to the corresponding sample means of the variance estimates. In Table 3, for the same cases we present the frequencies with which the approximate pivotals $X_0 = (\hat{\beta}_0 - \beta_0)/\hat{\sigma}_0$, $X_1 = (\hat{\beta}_1 - \beta_1)/\hat{\sigma}_1$, $X_2 = (\hat{\delta} - \delta)/\hat{\sigma}_2$ are less than or equal to the standard normal distribution 5, 25, 50, 75 and 95th percentiles. Given the small number of samples (100) for each case, the frequencies are reasonably close to the nominal values 5, 25, 50, 75 and 95. In addition, Q–Q plots of the 100 standardised statistics for each case show the X's to be approximately normal. These results suggest that the asymptotic approximations are, for the cases considered, sufficiently accurate for practical purposes.

5. An example

We consider some automobile warranty data which give the times and mileages at first and subsequent failures in a particular system on the car. Warranty coverage on each car extended for 12 000 miles or for one year from the date of sale, whichever occurred first. We will examine the distribution of time to first failure. The data are from $M = 8394$ cars manufactured in one plant during a two month period. Warranty claims were recorded

Table 3. *Frequencies with which approximate pivotals X_0, X_1, X_2 are less than or equal to the quantiles of the standard normal distribution, $q(5)$, $q(25)$, $q(50)$, $q(75)$, $q(95)$*

		Case 1					Case 2				
		$q(5)$	$q(25)$	$q(50)$	$q(75)$	$q(95)$	$q(5)$	$q(25)$	$q(50)$	$q(75)$	$q(95)$
L_C	X_0	6	28	54	77	92	2	20	41	75	98
	X_1	5	20	48	81	98	3	30	53	77	96
	X_2	7	21	48	71	94	2	25	55	80	97
PS_f ($p^f=0{\cdot}10$)	X_0	4	26	53	77	93	2	25	48	77	93
	X_1	4	26	53	80	95	6	28	52	74	94
	X_2	6	22	45	75	93	3	20	51	74	97
PS_{s1} ($p^s=0{\cdot}10$)	X_0	6	33	54	77	93	2	23	44	72	97
	X_1	3	23	46	76	98	6	29	53	73	94
	X_2	8	22	47	70	94	2	28	51	75	97
PS_{s2} ($p^s=0{\cdot}10$)	X_0	5	31	52	73	94	3	24	55	76	96
	X_1	4	26	51	82	99	3	22	46	76	93
	X_2	6	23	47	69	96	4	23	47	79	98

		Case 3				
		$q(5)$	$q(25)$	$q(50)$	$q(75)$	$q(95)$
L_C	X_0	2	28	50	80	96
	X_1	5	25	58	79	98
	X_2	2	20	48	75	99
PS_f ($p^f=0{\cdot}10$)	X_0	5	24	53	79	98
	X_1	5	29	54	82	97
	X_2	2	20	45	78	95
PS_{s1} ($p^s=0{\cdot}10$)	X_0	2	28	51	79	97
	X_1	6	25	51	81	97
	X_2	3	22	47	76	98
PS_{s2} ($p^s=0{\cdot}10$)	X_0	3	27	49	81	98
	X_1	6	25	54	81	97
	X_2	2	17	49	73	98

Case 1, $(\beta_0, \beta_1, \delta)=(-8{\cdot}6, 1{\cdot}0, 1{\cdot}0)$; Case 2, $(\beta_0, \beta_1, \delta)=(-15{\cdot}8, 1{\cdot}0, 2{\cdot}5)$; Case 3, $(\beta_0, \beta_1, \delta)=(-17{\cdot}6, 1{\cdot}0, 2{\cdot}5)$.
$X_0=(\hat\beta_0-\beta_0)/\tilde\sigma_0$, $X_1=(\hat\beta_1-\beta_1)/\tilde\sigma_1$, $X_2=(\hat\delta-\delta)/\tilde\sigma_2$, where $\tilde\sigma=\widehat{\mathrm{var}}^{1/2}$.

over the calendar period $(0, D]$, where 0 is by convention the day the first of the cars was sold and $D=547$ days. There were $m=823$ cars which experienced a failure, i.e. warranty claim, during the observation period. Let t_i represent the time of the first failure from car i, where 'time' can be either age, i.e. years since the car was sold, or mileage of the car. We know t_i if the failure occurs under warranty; otherwise all we know is the date of sale of the car and that there was no claim from it. In the following, we consider the situation where time is age, measured in years. The case where time is mileage could be similarly discussed.

It is likely that failure time depends on the rate at which mileage accumulates on a car. To investigate this and to deal with censoring times that are affected by mileage, we assume that the mileage of car i at age a is $w_i(a)=u_i a$, where u_i is the mileage accumulation rate in miles per year. This is an approximation since mileage does not accumulate exactly linearly, but has been found accurate enough for practical purposes. If car i experiences a first warranty claim at age a_i and mileage w_i then $t_i=a_i$ and we assume that $u_i=w_i/a_i$.

Then τ_i, the censoring time for car i, can be obtained as $\min\{(D-d_i)/365, A_0, W_0/u_i\}$, where d_i is the day car i was sold, and A_0 and W_0 are the age and mileage limits of the warranty plan, here one year and 12 000 miles, respectively. If car i does not experience a claim then all we know is that $t_i > \tau_i$; we do not know either u_i or τ_i. Diagnostics discussed below suggest it is suitable to suppose t_i $(i=1, \ldots, M)$ are independent with a Weibull distribution. To study the effect of mileage accumulation rate, we take u_i as a covariate, and adopt the Weibull proportional hazard regression model

$$f(t \mid u_i; \beta_0, \beta_1, \delta) = \delta t^{\delta-1} e^{\beta_0+\beta_1 u_i} \exp(-t^{\delta} e^{\beta_0+\beta_1 u_i}) \quad (i=1, \ldots, M).$$

It is reasonable to assume that t_i and τ_i are independent, given u_i. The units of t and u in the discussion below are years and thousands of miles per year, respectively.

Table 4. *Parameter estimates and standard errors from the automobile warranty data*

Methods	$\hat{\beta}_0$	$(\hat{\sigma}_0)$	$\hat{\beta}_1$	$(\hat{\sigma}_1)$	$\hat{\delta}$	$(\hat{\sigma}_2)$
L_T	1·8186	(0·4106)	−0·3298	(0·0874)	1·2674	(0·0501)
PS	−3·3014	(0·1315)	0·0947	(0·0097)	1·1859	(0·0409)

$\hat{\sigma}_0 = \{\hat{\mathrm{var}}(\hat{\beta}_0)\}^{1/2}$, $\hat{\sigma}_1 = \{\hat{\mathrm{var}}(\hat{\beta}_1)\}^{1/2}$, $\hat{\sigma}_2 = \{\hat{\mathrm{var}}(\hat{\delta})\}^{1/2}$.

The truncated data likelihood L_T given by (2) can be used to estimate the parameters; this utilises only the failure and truncation times (t_i, τ_i) for the m cars experiencing a failure. Table 4 shows parameter estimates and standard errors based on L_T. It is noted that the estimate of β_1 based on L_T is negative, implying that as the mileage accumulation rate increases the hazard function decreases. This result contradicts common sense, and is discussed further below. Fortunately, there is also some supplementary information about the population of cars. A customer survey sample of 607 cars of the same type and approximate geographical location as those in the warranty data base was taken, and among the information was the mileage accumulated at one year for each car. We used these data to construct an estimate of $G(\tau, u)$, the distribution of truncation time and mileage accumulation rate, as follows. The survey sample provided 607 mileage rates u_k $(k=1, \ldots, 607)$, which we used to estimate the marginal distribution of U: $\tilde{G}_U(u) = \sum I(u_k \leqslant u)/607$. We then took a random sample of size 607 from the sales dates d_i of the $M=8394$ cars in the population and estimated the distribution of $y_i = \min\{(D-d_i)/365, 1\}$ by $\tilde{G}_Y(y) = \sum I(y_k \leqslant y)/607$. Sales date and mileage accumulation rate can reasonably be assumed independent, so we then estimated $G(\tau, u)$ as $\tilde{G}(\tau, u)$, based on the fact that $\tau_i = \min\{(D-d_i)/365, 1, W_0/u_i\}$ with $W_0 = 12$, and $\bar{G}(\tau, u) = \bar{G}_Y(\tau)\{G_U(W_0/\tau) - G_U(u)\}$. The $\tilde{W}(\theta)$ in (4) was obtained as

$$\tilde{W}(\beta_0, \beta_1, \delta) = M \sum (-1, -u_k, -\delta/\tau_k)' \tau_k^{\delta} \exp(\beta_0 + \beta_1 u_k)/607,$$

with $\tau_k = \min(y_k, W_0/u_k)$.

We employed estimating function (4), treating $\{(y_k, u_k): k=1, \ldots, 607\}$ as a survey sample independent of the $M=8394$ cars giving the failure data. The dependence between the warranty data and the estimates $\tilde{G}(\tau, u)$ and $\tilde{W}(\beta_0, \beta_1, \delta)$ is slight, and standard errors for the estimates should be a reasonable reflection of precision. Table 4 shows the parameter estimates and standard errors. The estimates of β_0 and β_1 disagree markedly with the estimates provided by the truncated data alone (L_T). An examination of the estimates and the likelihood L_T indicates the source of the disagreement, and shows the implausibility of the truncated data estimates, as follows.

The estimate $\bar{F}(t\,|\,u)=\exp(-e^{\hat\beta_0+\hat\beta_1 u}t^{\hat\delta})$ based on $PS\{\theta;\,\tilde{W}(\theta)\}$ gives for $t=1$ and $u=12$ the probability 0·891 of no failure. This is in broad agreement with the warranty data, which gave 823 failures under warranty out of 8394 cars for which the average mileage accumulation rate is around 14 000 miles per year. For the estimate based on L_T the values $t=1$ and $u=12$ give essentially the same probability, 0·889. However, as noted above, the negative estimate of β_1 gives the unreasonable result that failure probabilities decrease as u increases. For the cars in this population most of the mileage accumulation rates are in the range 6–24. Figure 1 shows the estimates of the conditional cumulative distribution functions $F(t\,|\,u)/F(1\,|\,u)$ based on the estimates from L_T and $PS\{\theta;\,\tilde{W}(\theta)\}$, for $u=6$, 12, 24; it is seen that the graphs in each plot are close. Goodness-of-fit diagnostics based on approximately uniform (0, 1) residuals $\hat{F}(t_i\,|\,u_i)/\hat{F}(\tau_i\,|\,u_i)$ indicate that the Weibull model is reasonable, and the two estimates of $F(t\,|\,u)/F(1\,|\,u)$ are in good agreement with the observed warranty data. On the other hand, estimates of the unconditional distributions $F(t\,|\,u)$ based on $PS\{\theta;\,\tilde{W}(\theta)\}$ are very different from those based on L_T, except for accumulation rates around $u=12$. Figure 2 shows the estimates and approximate 95% pointwise confidence intervals of $F(t\,|\,u)$ for $0\leqslant t\leqslant 10$ for $u=6, 12, 24$ based on L_T and $PS\{\theta;\,\tilde{W}(\theta)\}$. Note that for Fig. 2(b) the upper confidence limit based on $PS\{\theta;\,\tilde{W}(\theta)\}$ is hard to distinguish from the estimate based on L_T. Whereas the estimates from $PS\{\theta;\,\tilde{W}(\theta)\}$ are plausible, those from L_T are not.

Further insight may be gained by considering the parameter $\psi=\beta_0+12\beta_1$; the estimate of ψ from L_T is very close to that from $PS\{\theta\colon\,\tilde{W}(\theta)\}$, relative to their standard errors. The estimates of δ are also similar, with the result that unconditional estimates of the distribution function near $u=12$ are similar, as well as the conditional estimates shown in Fig. 1. The line $\beta_0+12\beta_1=\hat\phi$, where $\hat\phi=\hat\beta_0+12\hat\beta_1$ is the estimate based on PS, passes through the middle of an approximate 95% confidence region for (β_0,β_1) based on $\log(L_T)$. However, this region suggests quite different values for (β_0,β_1) than do $\hat\beta_0, \hat\beta_1$ from PS.

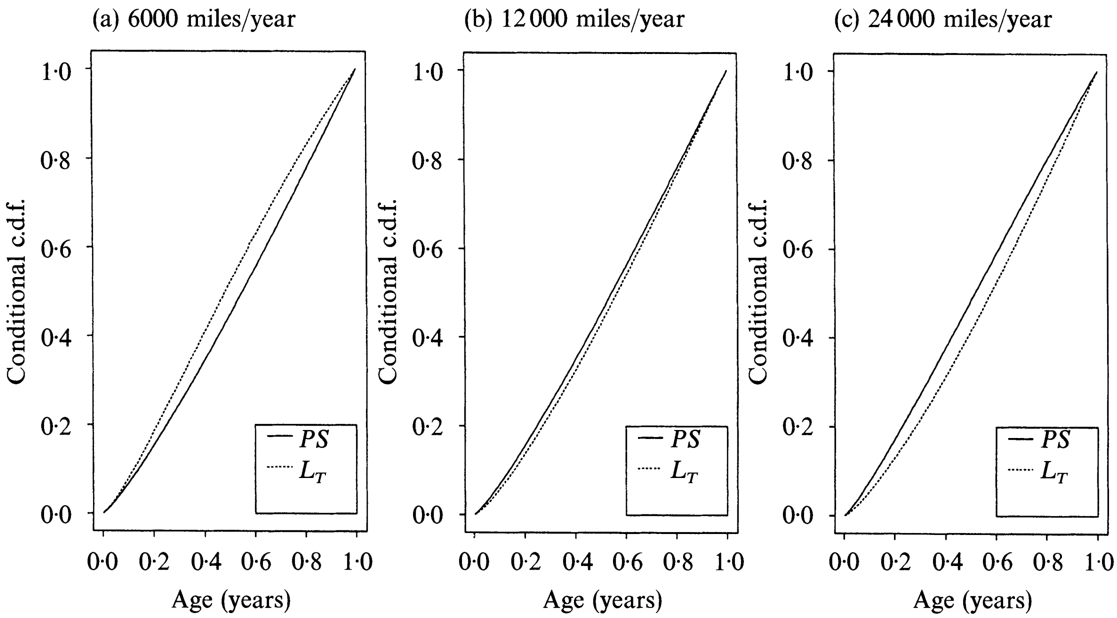

Fig. 1. Estimates of the cumulative failure distribution, conditional on failure within one year, from L_T and $PS\{\theta;\,\tilde{W}(\theta)\}$, for usage rates of 6000, 12 000 and 24 000 miles/year.

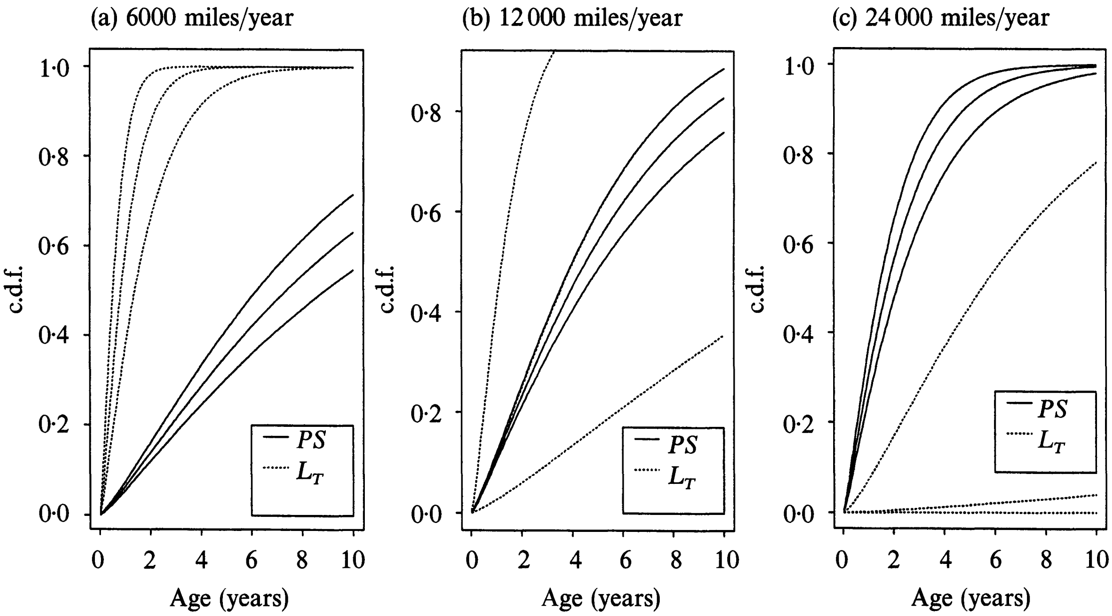

Fig. 2. Estimates of the cumulative failure distribution, for usage rates of 6000, 12 000 and 24 000 miles/year.

In the current situation the car manufacturer would like to estimate failure probabilities beyond one year, and we have shown such estimates in Fig. 2. The procedure based on $PS\{\theta;\ \tilde{W}(\theta)\}$ gives plausible estimates, but the warranty data do not of course allow us to check the adequacy of the Weibull model beyond one year. For that we must consult other sources of information. The estimates in Fig. 2 must consequently be considered tentative for ages beyond one year.

6. Concluding remarks

The methods in the paper may be extended to deal with more complex supplementary sampling schemes and with situations where censoring times, but not covariates, are known for non-responding individuals. The approach in §§ 2 and 3 may also be used with estimating functions which are not based on a likelihood. It is also possible to construct alternative estimating functions and, in some cases, semiparametric likelihoods. For example, the methods presented in § 2 do not require direct estimation of $G(\tau, z)$; other methods use the supplementary data to estimate G and hence to estimate $\mathrm{pr}(T_i > \mathcal{T}_i)$ for $i \in \mathcal{O}_2$. Hu and Lawless discuss a variety of approaches in an unpublished technical report. Some approaches, see also Carroll & Wand (1991), Pepe & Fleming (1991) and Pepe (1992), may be characterised in terms of an estimation problem where $U(Y;\ \theta, G)$ is an unbiased estimating function for θ based on observed data Y and an unknown distribution function G which is a nuisance parameter. The approach taken is to estimate G empirically by utilising supplementary information additional to Y. Another type of approach is to use supplementary data to estimate the 'incomplete' part of a complete data log-likelihood or score, such as $S_C(\theta)$ in this paper; see also Kalbfleisch & Lawless (1988a, b), Wild (1991) and Reilly & Pepe (1995). Specific instances of these approaches require study in order to deal with regularity conditions for asymptotics, small sample behaviour of estimators, details regarding the supplementary observation and so on. Some general treatment is possible, and will be discussed elsewhere.

The simulations and example in this paper dealt with situations where neither the expected number of failures nor the supplementary sampling fraction is too small. The variance estimates and approximate pivotals on which tests are based perform well here, but in applications where either is small it would be wise to carry out a small simulation study. Finally, for the methods in this paper to be effective the supplementary sample must be representative of the population of individuals generating the response times and censoring times. In the case of a follow-up sample selected from nonresponding individuals in the population this condition is satisfied, but for applications like that in § 5 caution is called for.

Acknowledgement

This research was supported in part by grants to J. F. Lawless from the Natural Sciences and Engineering Research Council of Canada and General Motors Canada. The authors thank R. J. MacKay and two anonymous reviewers for helpful remarks.

Appendix 1

Proof of asymptotic results

Let

$$\Psi(t, \tau, z \mid \theta) = I(t \leqslant \tau) \log f(t \mid \theta; \tau, z) + I(t > \tau) \log \bar{F}(\tau \mid \theta; \tau, z).$$

Suppose that:

(i) θ_0 is the true value of the parameter, where $\theta_0 \in A$, an open subset of Θ, and that $0 < P_\theta(T > \mathscr{T}) < 1$ if $\theta \in A$;

(ii) the usual regularity conditions (Serfling, 1980, Ch. 4) hold for both $f(t \mid \theta; \tau, z)$ and $\bar{F}(\tau \mid \theta; \tau, z)$;

(iii) for all $t > 0$ and $(\tau, z) \in \mathscr{A}$, Ψ_θ, $\Psi_{\theta\theta}$ and $\Psi_{\theta\theta\theta}$, the first, second and third order partial derivatives of Ψ, exist for all $\theta \in A$; $\Sigma(\theta_0) := E_{\theta_0} \Psi_\theta \Psi'_\theta$ exists, and is positive definitive, for all $\theta \in A$;

(iv) each element of $\Psi_{\theta\theta\theta}$ is bounded by an integrable function for all $\theta \in A$; $|\Psi_{\theta\theta\theta}| \leqslant H(t, \tau, z)$, where $E\{H(T, \mathscr{T}, Z)\}$ exists.

If $\tilde{W}(\theta)$ is an estimate of $W(\theta)$ in (3) and $\Delta(\theta) = \tilde{W}(\theta) - W(\theta)$, also assume that

(v) as $M \to \infty$, uniformly $M^{-1} \partial^l \Delta(\theta) / \partial \theta^l \to 0$ almost surely, with $l = 0, 1, 2$, and $\Delta(\theta)/\sqrt{M} \to N\{0, K(\theta_0)\}$, in distribution.

Proof of Theorem 1. Let $\theta \in A$ be fixed. We have by Taylor expansion of $PS(\theta) := PS\{\theta; \tilde{W}(\theta)\}$ in (4), about the point θ_0:

$$PS(\theta) = PS(\theta_0) + \left.\frac{\partial PS(\lambda)}{\partial \lambda}\right|_{\lambda = \theta_0} (\theta - \theta_0) + \frac{1}{2} (\theta - \theta_0)' \left.\frac{\partial^2 PS(\lambda)}{\partial \lambda^2}\right|_{\lambda = \xi} (\theta - \theta_0), \tag{A1}$$

where $\xi \in A$ and $\| \xi - \theta_0 \| \leqslant \| \theta - \theta_0 \|$. By recalling the properties of the score function $S_C(\theta)$ of (3), and noting the conditions above, we know that, as $M \to \infty$,

$$\frac{1}{M} PS(\theta_0) \to 0, \quad \frac{1}{M} \left.\frac{\partial PS(\lambda)}{\partial \lambda}\right|_{\lambda = \theta_0} \to -\Sigma(\theta_0),$$

$$\frac{1}{M} \left| \left.\frac{\partial^2 PS(\lambda)}{\partial \lambda^2}\right|_{\lambda = \xi} \right| \leqslant E\{H(T, \mathscr{T}, Z)\} + J \tag{A2}$$

almost surely, when $\|\theta-\theta_0\|$ is small enough, where J is a matrix of ones with the same dimensions as $H(t,\tau,z)$. Now let $\varepsilon>0$ be given, such that $\theta_1=\theta_0-\varepsilon 1$, and $\theta_2=\theta_0+\varepsilon 1$ lie in A, (A2) holds, and

$$\varepsilon<\frac{\|1'[E\{H(T,\mathscr{T},Z)\}+J]1\|}{2\|\Sigma(\theta_0)1\|},$$

where 1 is a $p\times 1$ vector of 1's. Then, by (A1), we see that both

$$\|M^{-1}PS(\theta_1)-\varepsilon\Sigma(\theta_0)1\|,\quad \|M^{-1}PS(\theta_2)+\varepsilon\Sigma(\theta_0)1\|$$

are bounded almost surely by $3\varepsilon\|\Sigma(\theta_0)1\|/4$, if M is sufficiently large. For such M, the interval

$$\left[\frac{1}{M}PS(\theta_2),\frac{1}{M}PS(\theta_1)\right]$$

contains the point 0 almost surely and hence, by the continuity of $PS(\theta)$ with respect to θ, the interval $[\theta_0-\varepsilon 1,\theta_0+\varepsilon 1]$ contains a solution of $PS\{\theta;\,\tilde{W}(\theta)\}=0$ almost surely. We have thus proved the strong consistency of the estimator $\hat{\theta}$.

Let θ in (A1) be $\hat{\theta}$. Then,

$$PS(\theta_0)=-\left\{\left.\frac{\partial PS(\lambda)}{\partial\lambda}\right|_{\lambda=\theta_0}+\frac{1}{2}(\hat{\theta}-\theta_0)'\left.\frac{\partial^2 PS(\lambda)}{\partial\lambda^2}\right|_{\lambda=\xi}\right\}(\hat{\theta}-\theta_0).$$

Since almost surely

$$\frac{1}{M}\left\{\left.\frac{\partial PS(\lambda)}{\partial\lambda}\right|_{\lambda=\theta_0}+\frac{1}{2}(\hat{\theta}-\theta_0)'\left.\frac{\partial^2 PS(\lambda)}{\partial\lambda^2}\right|_{\lambda=\xi}\right\}\to-\Sigma(\theta_0),$$

we can prove $\sqrt{M}(\hat{\theta}-\theta_0)$ is asymptotically normal by noting the conditions above and the asymptotic normality of $S_C(\theta)$. Notice that the asymptotic covariance of $PS(\theta_0)=S_C(\theta_0)+\Delta(\theta_0)$ is $\Sigma(\theta_0)+2H(\theta_0)+K(\theta_0)$. This gives the asymptotic covariance matrix of $\sqrt{M}(\hat{\theta}-\theta_0)$ from (9). □

APPENDIX 2

Implementation with supplementary samples

Supplementary follow-up data. In this case, we see that

$$\Delta_f(\theta)=\tilde{W}_f(\theta)-W(\theta)=\sum_{i\in\mathscr{P}}\left(\frac{R_{i2}}{p^f}-1\right)(1-R_{i1})v_i(\theta),$$

a sum of independent identically distributed random variables. It is easy to check that condition (v) is satisfied if $p^f>0$. Noting the sampling scheme, we know $E\{\Delta_f(\theta)\,|\,\mathscr{O}\}=0$, where $\mathscr{O}=\{(\tau_i,z_i):i\in\mathscr{P}\}$, and, then, $H(\theta_0)=0$;

$$\operatorname{var}\{\Delta_f(\theta)\,|\,\mathscr{O}\}=(M-m)(1/p^f-1)\operatorname{var}\{V(\theta)\,|\,\mathscr{O}\},$$

where $V(\theta)=\partial\log\bar{F}(\mathscr{T}\,|\,\theta;\,\mathscr{T},Z)/\partial\theta$. Thus, the consistent estimator $\hat{K}_f$ from (11) for $K(\theta_0)$ is obtained.

Supplementary random sample from $\mathscr{P}$. Notice that

$$\Delta_{s1}(\theta)=\sum_{i\in\mathscr{P}}\left(\frac{Q_i}{p^s}-1\right)v_i(\theta),$$

where $Q_i=I(i\in\mathscr{S}^s)$. Thus condition (v) is satisfied if $p^s>0$. Since

$$E(Q_i\,|\,\mathscr{O})=p^s,\quad \operatorname{var}(Q_i\,|\,\mathscr{O})=p^s(1-p^s),$$

we know that $H(\theta_0) = 0$, and $K(\theta_0) = (1/p^s - 1)\,\mathrm{var}\{V(\theta_0)\}$, which leads to the consistent estimator $\hat{K}_{s1}$ from (12).

Supplementary random sample independent of $\mathscr{P}$. We know that

$$\Delta_{s2}(\theta) = \frac{1}{p^s} \sum_{j \in \mathscr{S}^s} v_j(\theta) + \sum_{i \in \mathscr{P}} v_i(\theta),$$

so condition (v) holds provided $p^s > 0$. Based on

$$K(\theta_0) = (1/p^s + 1)\,\mathrm{var}\{V(\theta_0)\}, \quad H(\theta_0) = -\mathrm{cov}\{\delta U(\theta_0), V(\theta)\} - \mathrm{var}\{V(\theta_0)\},$$

where δ is the indicator of $\mathscr{P}^*$ and $U(\theta) = \partial \log\{f(T|\theta; \mathscr{T}, Z)/\bar{F}(\mathscr{T}|\theta; \mathscr{T}, Z)\}/\partial\theta$, we obtain the consistent estimators $\hat{K}_{s2}$ from (13), and $\hat{H}_{s2}$ from (15).

References

Carroll, R. J. & Wand, M. P. (1991). Semiparametric estimation in logistic measurement error models. *J. R. Statist. Soc.* B **53**, 573–85.

Hsieh, D. A., Manski, C. F. & McFadden, O. (1985). Estimation of response probabilities from augmented retrospective observations. *J. Am. Statist. Assoc.* **80**, 651–62.

Hu, X. J. & Lawless, J. F. (1996). Estimation of rate and mean functions from truncated recurrent event data. *J. Am. Statist. Assoc.* **91**, 300–10.

Jewell, N. P. (1985). Least squares regression with data arising from stratified samples of the dependent variable. *Biometrika* **72**, 11–21.

Kalbfleisch, J. D. & Lawless, J. F. (1988a). Estimation of reliability from field performance studies (with Discussion). *Technometrics* **30**, 365–88.

Kalbfleisch, J. D. & Lawless, J. F. (1988b). Likelihood analysis of multi-state models for disease incidence and mortality. *Statist. Med.* **7**, 149–60.

Kalbfleisch, J. D. & Lawless, J. F. (1989). Inference based on retrospective ascertainments: an analysis of the data on transfusion-related AIDS, *J. Am. Statist. Assoc.* **84**, 360–72.

Pepe, M. S. (1992). Inference using surrogate outcome data and a validation sample. *Biometrika* **79**, 355–65.

Pepe, M. S. & Fleming, T. R. (1991). A nonparametric method for dealing with mis-measured covariate data. *J. Am. Statist. Assoc.* **86**, 108–13.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.

Reilly, M. & Pepe, M. S. (1995). A mean score method for missing and auxiliary covariate data in regression models. *Biometrika* **82**, 299–314.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley.

Suzuki, K. (1985a). Estimation method of lifetime based on the record of failures during the warranty period. *J. Am. Statist. Assoc.* **80**, 68–72.

Suzuki, K. (1985b). Nonparametric estimation of lifetime distribution from incomplete field data. *Technometrics* **27**, 263–71.

Suzuki, K. (1987). Analysis of field failure data from a nonhomogeneous Poisson process. *Rep. Statist. Appl. Res. Union Jap. Sci. Eng.* **34**, 6–15.

Suzuki, K. (1993). Estimation of lifetime distribution using the relationship of calendar time and usage time. *Rep. Statist. Appl. Res. Union Jap. Sci. Eng.* **40**, 10–22.

Suzuki, K. (1995). Role of field performance data and its analysis. In *Recent Advances in Life-Testing and Reliability*, Ed. N. Balakrishnan, pp. 141–51. Boca Raton: CRC Press.

Wild, C. J. (1991). Fitting prospective regression models to case-control data. *Biometrika* **78**, 705–17.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163–77.

[*Received April* 1995. *Revised January* 1996]