Estimation with clustered censored survival data with missing covariates in the marginal Cox model

Michael Parzen\textsuperscript{1}, Stuart Lipsitz\textsuperscript{2}, Amy Herring\textsuperscript{3}, and Joseph G. Ibrahim\textsuperscript{3}

(1) Emory University

(2) Harvard Medical School

(3) University of North Carolina

Submitted to the \textit{Journal of the American Statistical Association}
SUMMARY

Studies in the health sciences often give rise to correlated, possibly censored, survival data. Wei et al. (1989) and Lee, et al. (1992) show that, if the marginal distributions of the correlated survival times follow a proportional hazards model, then the estimates from Cox’s partial likelihood (1972), naively treating the correlated survival times as independent, give consistent parameter estimates. Here, observations within a cluster can have some missing covariate values. With non-clustered data, Herring and Ibrahim (2001) propose a set of estimating equations to consistently estimate the parameters of Cox’s proportional hazards model when some covariate values are missing. These estimating equations can be solved by an algorithm similar to the EM algorithm. We show in this paper that if one naively treats observations within a cluster as independent, that one can still use the estimating equations of Herring and Ibrahim (2001) to obtain consistent estimates of the relative risk parameters. This method requires the estimation of the parameters of the distribution of the covariates, and allows the missing data mechanism for an observation in a cluster to depend on observed data for the observation. We present results from a clinical trial (Falkson et al., 1990) with five covariates, four of which have some missing values. In the trial, the clusters are the hospitals in which the patients were treated.

Key words: Ignorable missing data, Missing at random, Monte-Carlo methods, Non-informative censoring, score vector.
1 Introduction

Clustered survival data arise often in biomedical studies; observations within the cluster tend to be correlated. In this paper, the data consists of $n$ independent clusters; the response vector for each cluster is a vector of correlated, possibly censored, survival times. The example in this paper is from two Eastern Cooperative Oncology Group (ECOG) clinical trials to evaluate patients with primary liver cancer (Falkson et al. 1990; Falkson et al., 1994). The primary interest here is how the outcome survival, time from entry on the study until death, differs with respect to five dichotomous baseline covariates. The five covariates are age, categorized as less than 60 or greater than or equal to 60; associated Jaundice, yes or no; associated Hepatitis, yes or no; and two biochemical markers, Alpha fetoprotein and Anti Hepatitis B antigen, each classified as normal or abnormal.

In many clinical trials, including those undertaken by large cooperative cancer groups (CALGB, ECOG, SWOG, EORTC), patients enter the trial from an institution (i.e. hospital). Not infrequently, patients from the same institution have similar outcomes (i.e. outcomes which are correlated), due, possibly, to unmeasured variables such as the skill or training of the staff or the quality of the hospital equipment. Institutions can then be thought of as clusters, and patients as the units within these clusters. In the liver cancer study, the data include 182 observations from $n = 31$ institutions. Table 1 shows the data from 8 of these clusters.

In Table 1, we see that there are missing covariate data. Table 2 shows the percent missing (over the 182 observations) on each covariate. Note that 46 (25%) of the 182 observations are missing at least one variable. The biochemical markers require blood to be drawn and possibly complicated laboratory tests, and thus have the highest proportion missing. We allow for right censored observations and assume noninformative censoring.
for an observation within the cluster. We assume, for a given member of a cluster, that
the probability that data are missing only depends on the observed data for that member
(observed covariates, censoring time, and censoring indicator), but no data on from other
members of the cluster, or members of other clusters. When the members of the clusters
are patients, as in this study, it is realistic to assume that missingness for one patient is
independent of missingness for a different patient. For the liver cancer study discussed
here, since there are so few missing values of age and associated hepatitis, it is reasonable
to assume that these data are missing completely at random. Further, as discussed in
Section 6, it is plausible that missingness in the biochemical markers can depend on the
observed data for a subject. As we will see in section 5, a patient with jaundice is much
more likely to have the other covariates observed, possibly due to the fact that doctors
of patients with jaundice may want to see all of the laboratory tests and biochemical
markers at baseline. Also, patients who live longer (or have a longer censoring time) are
much more likely to have the other covariates observed, possibly due to the fact that the
data managers have more time to query the laboratories for the values of the baseline
biochemical markers; it may not be worth the trouble to update the file of a patient who
has already passed away. Further, doctors may order follow-up laboratory tests for a
patient who lives longer, and thus the data manager will query the laboratory for the
baseline value in order to look at the change from baseline.

To estimate relative risks with correlated survival data with no missing data, Wei et
al. (1989) and Lee, et al. (1992) propose the use of the Cox proportional hazards model
for an observation within a cluster. To consistently estimate the regression parameters,
they suggest naively treating the observations within the cluster as independent and using
the Cox partial likelihood. However, because of the correlation between survival times in
a cluster, the inverse of the information matrix may not be a consistent estimate of the
asymptotic variance. Wei, et al. (1989) and Lee, et al. (1992) propose a robust variance estimate that is consistent for the asymptotic variance of the estimates from the Cox model.

With missing covariates, one possibility would be to discard the observations in the cluster with missing covariate data, and apply the Wei et al. (1989) approach. However, this can lead to highly biased and inefficient estimates if the missing data fraction is large and the missing data are not missing at completely at random (MCAR) as defined by Rubin (1976). Thus, it is desirable in these situations to estimate the parameters using all the data if possible. If the investigator is willing to assume a parametric model for the marginal distributions, Lipsitz and Ibrahim (2000) show that if one naively treats observations within a cluster as independent, that one can still use the maximum likelihood estimating equations via the EM-algorithm to obtain consistent estimates of the relative risk parameters. However, we assume the marginal model follows the Cox proportional hazards model, so that the Lipsitz and Ibrahim (2000) method does not apply.

With missing covariate data in univariate survival data following the Cox model, Herring and Ibrahim (2001) proposed a set of estimating equations which can be solved by an algorithm similar to the EM algorithm. These estimating equations produce consistent estimates as long as the data are missing at random (as discussed earlier). i.e., missingness only depends on observed data for that subject. In this paper, we show that these estimating equations can be used to consistently estimate the marginal Cox regression parameters. In particular, one can naively treat the observations within the cluster as independent and use these estimating equations to obtain a consistent estimates of the relative risk parameters. Although the regression parameter estimates are consistent when naively assuming the observations in a cluster as independent, because of the correlation between survival times in a cluster, the inverse of the information matrix may not be a
consistent estimate of the asymptotic variance. To consistently estimate the variance of
the estimates, one needs to use a robust variance estimate such as a ‘sandwich’ estimator
(White, 1982).

2 Distributions

Let \( T_{ik} \) be the failure time for the \( k^{th} \) member of cluster \( i \), \( i = 1, ..., n; \ k = 1, ..., n_i \), and \( z_{ik} = [z_{ik1}, ..., z_{ikP}]' \) be a \((P \times 1)\) vector of covariates. If \( T_{ik} \) follows the Cox proportional
hazards regression model (1972), then the hazard function for \( T_{ik} \) at time \( t > 0 \), conditional
on \( z_{ik} \) is

\[
\lambda(t|z_{ik}) = \lambda_0(t) \exp(\beta'z_{ik}),
\]

where \( \lambda_0(t) \) is an arbitrary baseline hazard function, and \( \beta \) is a \( P \times 1 \) vector of regression
coefficients.

Using (1), the density for \( T_{ik} \) can be written as

\[
p(t_{ik}|z_{ik}; \beta) = \lambda(t_{ik}|z_{ik}; \beta)S(t_{ik}|z_{ik}; \beta)
\]

\[
= \lambda_0(t_{ik}) \exp(\beta'z_{ik}) \exp \left\{ -e^{\beta'z_{ik} \Lambda_0(t_{ik})} \right\},
\]

where

\[
S(t_{ik}|z_{ik}; \beta) = \text{pr}(T_{ik} > t_{ik}|z_{ik}, \beta) = \exp \left\{ -e^{\beta'z_{ik} \Lambda(t_{ik})} \right\}
\]
is the survivor function for \( T_{ik} \), and

\[
\Lambda_0(t) = \int_0^t \lambda_0(u)du
\]
is the cumulative baseline hazard function.

The survival time is often right censored, so we let \( U_{ik} \) be the censoring time. Thus, instead of \( T_{ik} \), we observe \( X_{ik} = \min(T_{ik}, U_{ik}) \) where the censoring indicator \( \delta_{ik} = I[T_{ik} \leq U_{ik}] \) equals 1 if \( Y_{ik} \) is a failure time and 0 if it is right censored. Conditional on the
covariates and assuming non-informative censoring (see Lawless, 1982), the probability distribution for \((x_{ik}, \delta_{ik}|z_{ik})\) is proportional to

\[
p(x_{ik}, \delta_{ik}|z_{ik}, \beta) = p(x_{ik}|z_{ik}, \beta)^{\delta_{ik}} S(x_{ik}|z_{ik}, \beta)^{1-\delta_{ik}}.
\]

\[
= \lambda(x_{ik}|z_{ik}, \beta)^{\delta_{ik}} S(x_{ik}|z_{ik}, \beta)
\]

\[
= [\lambda_0(x_{ik}) \exp(\beta' z_{ik})]^\delta_{ik} \exp \left\{ -e^{\beta' z_{ik}} \Lambda_0(x_{ik}) \right\}.
\]

Thus, for the \(k^{th}\) member of cluster \(i\), the log-likelihood for \(\beta\) is given by

\[
\ell(\beta; x_{ik}, \delta_{ik}, z_{ik}) = \delta_{ik} \log[\lambda_0(x_{ik})] + \delta_{ik}(\beta' z_{ik}) - e^{\beta' z_{ik}} \Lambda_0(x_{ik}).
\]

Equation (4) contains the nuisance parameters \(\lambda_0(t)\), and \(\Lambda_0(t)\), which Cox (1972) eliminated using a partial likelihood.

Since some elements of \(z_{ik}\) can be missing, we must also consider its distribution. Let \(z_{ik}\) have density \(p(z_{ik}|\alpha)\) indexed by \(\alpha\), where \(\alpha\) is distinct from \(\beta\) and \(\lambda_0(t)\). With missing elements of \(z_{ik}\), it is convenient to introduce a \((P \times 1)\) random vector for the \(k^{th}\) member of cluster \(i\), \(R_{ik}\), whose \(p^{th}\) component, \(R_{ikp}\), equals 1 if the \(p^{th}\) component of \(z_{ik}\) is observed, and equals 0 if it is missing. The distribution of \(R_{ik}\) given \((x_{ik}, \delta_{ik}, x_{ik})\) is called the "missing data mechanism", and has a multinomial distribution with \(2^P\) cell probabilities,

\[
p(r_{ik}|x_{ik}, \delta_{ik}, z_{ik}, \omega) = \Pr\{R_{ik1} = r_1, \ldots, R_{ikP} = r_p|x_{ik}, \delta_{ik}, z_{ik}, \omega\},
\]

parameterized by \(\omega\). In this paper, we assume that (5) depends on \((x_{ik}, \delta_{ik})\), and the observed components of \(z_{ik}\), say \(z_{obs,ik}\). In general, we write \(z_{ik} = (z_{mis,ik}, z_{obs,ik})\) where \(z_{mis,ik}\) is the missing component of \(z_{ik}\). Thus, we assume

\[
p(r_{ik}|x_{ik}, \delta_{ik}, z_{ik}, \omega) = p(r_{ik}|x_{ik}, \delta_{ik}, z_{obs,ik}, \omega).
\]

We assume that given \((x_{ik}, \delta_{ik}, z_{obs,ik})\), \(r_{ik}\) is independent of data for any other member of cluster \(i\), or data from any other cluster. For the liver cancer study, it is plausible that
missingness will depend on \((x_{ik}, \delta_{ik})\) (and not necessarily the underlying \(T_{ik}\)). As we will see in section 5, patients who have a longer censoring time are much more likely to have the other covariates observed, possibly due to the fact that the data managers have more time to query the laboratories for the values of the baseline biochemical markers; it may not be worth the trouble to update the file of a patient who has already passed away. Further, doctors may order follow-up laboratory tests for a patient who lives longer, and thus the data manager will query the laboratory for the baseline value in order to look at the change from baseline. If observations within a cluster are independent, (6) is a missing at random assumption (Little and Rubin, 1987). Our main interest is in the estimation of \(\beta\).

3 The estimating equations with no missing data

First, we describe the estimating equations naively assuming observations in a cluster are independent, and there are no missing data. Using counting process notation, the information in the data can be represented by

\[
\{N_{ik}(t), Y_{ik}(t), Z_{ikk} : 0 < t < \infty\},
\]

where \(N_{ik}(t)\) takes value 1 if the \(k^{th}\) member in cluster \(i\) has been observed to fail prior to time \(t\), and 0 otherwise, and \(Y_{ik}(t)\) takes value 1 if the \(k^{th}\) member in cluster \(i\) is at risk at time \(t\) and value 0 otherwise. Although observations within clusters tend to be correlated, for hazard (1), the Cox partial likelihood score vector obtained by naively assuming the observations within a cluster are independent, is

\[
u(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{n_i} \int_{0}^{\infty} \left\{Z_{ik} - \bar{Z}(s, \beta)\right\} dN_{ik}(s),
\]

(7)
where
\[
\bar{Z}(s, \beta) = \frac{\sum_{j=1}^{n} \sum_{k=1}^{n_{ik}} Z_{jk} Y_{jk}(s) e^{\beta z_{jk}}}{\sum_{j=1}^{n} \sum_{k=1}^{n} Y_{jk}(s) e^{\beta z_{jk}}}
\]
is a weighted average of the \(Z_{ik}\)'s for those at risk at time \(s\), and \(dN_{ik}(s) = N_{ik}(s) - N_{ik}(s^-)\) is a binary random variable which equals 1 if member \(k\) in cluster \(i\) fails at time \(s\), and equals 0 otherwise. The estimate \(\beta\) is the solution to \(u(\beta) = 0\). A consistent estimate of \(\Lambda_{0}(s)\) is obtained via
\[
\hat{\Lambda}_{0}(s) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{n_{i}} I[T_{ik} < s] \hat{\lambda}_{0}(t_{ik})}{\sum_{i=1}^{n} \sum_{k=1}^{n_{i}} \hat{\lambda}_{0}(t_{ik})}.
\]

Even though observations in a cluster are not independent, Wei et al. (1989) show that \(\beta\) is still consistent and asymptotically normal. In particular, Wei et al. (1989) show that \(u(\beta)\) has asymptotic mean 0 even though observations in a cluster can be correlated; thus, using method of moments theory (Casella & Berger, 1990, ch. 7), \(\hat{\beta}\) is consistent because \(u(\beta)\) has asymptotic mean 0 and we are solving \(u(\hat{\beta}) = 0\) for \(\hat{\beta}\). Note, for future reference, Wei et al. (1989) show that \(u(\beta)\) is asymptotically equivalent to the random variable
\[
u_{A}(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{n_{i}} \int_{0}^{\infty} \{z_{ik} - \mu(s, \beta)\} dN_{ik}(s), \tag{8}
\]
where \(\mu(s, \beta)\) is the asymptotic limit of \(\bar{Z}(s, \beta)\), i.e.,
\[
\bar{Z}(s, \beta) \xrightarrow{p} \lim_{n \to \infty} \frac{\sum_{j=1}^{n} \sum_{k=1}^{n_{ik}} E[Y_{jk}(s)Z_{jk} e^{\beta z_{jk}}]}{\sum_{j=1}^{n} \sum_{k=1}^{n} E[Y_{jk}(s) e^{\beta z_{jk}}]} = \mu(s, \beta); \tag{9}
\]
further, using the results of Wei et al. (1989),
\[
E[\nu_{A}(\beta)] = \sum_{i=1}^{n} \sum_{k=1}^{n_{i}} \int_{0}^{\infty} E \{[z_{ik} - \mu(s, \beta)] dN_{ik}(s)\} = 0. \tag{10}
\]
Now, we discuss the estimating equations when some of the components of $z_{ik}$ are missing.

Write $z_{ik} = (z_{mis,ik}, z_{obs,ik})$, where $z_{mis,ik}$ contains the missing components of $z_{ik}$, and $z_{obs,ik}$ contains the observed components of $z_{ik}$. We propose the estimating equations $s(\hat{\beta}, \hat{\lambda}) = 0$ to estimate $(\beta, \lambda)$, where

$$s(\beta, \lambda) = \begin{bmatrix} s_1(\beta, \lambda) \\ s_2(\beta, \lambda) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n \sum_{k=1}^n \int_0^\infty E[Z_{ik} | z_{obs,ik}, x_{ik}, \delta_{ik}] - \bar{Z}_E(s, \beta) dN_{ik}(s) \\ \sum_{i=1}^n \sum_{k=1}^n dN_{ik}(t) - \lambda_0(t) Y_{ik}(t) E[e^{\beta'z_{ik}} | z_{obs,ik}, x_{ik}, \delta_{ik}] \end{bmatrix},$$

(11)

where

$$\bar{Z}_E(s, \beta) = \frac{\sum_{j=1}^n \sum_{k=1}^{n_{ik}} Y_{jk}(s) E[Z_{jfk} e^{\beta'z_{jfk}} | z_{obs,ik}, x_{ik}, \delta_{ik}]}{\sum_{j=1}^n \sum_{k=1}^{n_{ik}} Y_{jk}(s) E[e^{\beta'z_{jfk}} | z_{obs,ik}, x_{ik}, \delta_{ik}]}$$

(12)

Using results similar to Herring and Ibrahim (2001), if the conditional expectations

$$E[Z_{ik} | z_{obs,ik}, x_{ik}, \delta_{ik}];$$

(13)

$$E[e^{\beta'z_{jfk}} | z_{obs,ik}, x_{ik}, \delta_{ik}];$$

(14)

and

$$E[Z_{jfk} e^{\beta'z_{jfk}} | z_{obs,ik}, x_{ik}, \delta_{ik}]$$

(15)

are correctly specified in (11) and (12), then the solution $(\hat{\beta}, \hat{\lambda})$ to $s(\hat{\beta}, \hat{\lambda}) = 0$ will be consistent and asymptotically normal.

Here, we heuristically argue that $\hat{\beta}$ is consistent. Using the results of Herring and Ibrahim (2001), as long as the conditional expectations in (13), (14), and (15) are correctly specified, as $n \to \infty$,

$$\bar{Z}_E(s, \beta) \xrightarrow{p} \lim_{n \to \infty} \frac{\sum_{j=1}^n \sum_{k=1}^n E[Y_{jk}(s) Z_{jfk} e^{\beta'z_{jfk}}]}{\sum_{j=1}^n \sum_{k=1}^n E[Y_{jk}(s) e^{\beta'z_{jfk}}]} = \mu(s, \beta),$$
where \( \mu(s, \beta) \) is identical to that given in (9). Then, \( s_1(\beta, \lambda) \) is asymptotically equivalent to the random variable

\[
s_{1A}(\beta, \lambda) = \sum_{i=1}^n \sum_{k=1}^n \int_0^\infty \{E[Z_{ik}|z_{obs,ik}, x_{ik}, \delta_{ik}] - \mu(s, \beta)\} dN_{ik}(s).
\]  (16)

Further, using conditional probability,

\[
E[Z_{ik}dN_{ik}(s)] = E\{dN_{ik}(s)E[Z_{ik}|z_{obs,ik}, x_{ik}, \delta_{ik}]\},
\]

so that

\[
E[s_{1A}(\beta, \lambda)] = \sum_{i=1}^n \sum_{k=1}^n \int_0^\infty E\{[Z_{ik} - \mu(s, \beta)]dN_{ik}(s)\} = E[u_A(\beta)] = 0,
\]  (17)

where \( u_A(\beta) \) is given in (10). Similarly, one can show that \( E[s_2(\beta, \lambda)] \) has asymptotic mean 0. Thus, since \( s(\beta, \lambda) \) has asymptotic mean 0, and we are solving \( s(\hat{\beta}, \hat{\lambda}) = 0 \), again using method of moment ideas, \((\hat{\beta}, \hat{\lambda})\) is consistent.

The consistency of \((\hat{\beta}, \hat{\lambda})\) does require the correct specification of the conditional expectations (13), (14), and (15). To correctly these conditional expectations, we must correctly specify and estimate the conditional distribution \( p(z_{mis,ik}|z_{obs,ik}, x_{ik}, \delta_{ik}) \). Assuming \( z_{mis,ik} \) is categorical,

\[
p(z_{mis,ik}|z_{obs,ik}, x_{ik}, \delta_{ik}) = \frac{p(x_{ik}, \delta_{ik}|z_{obs,ik}, x_{ik}, \delta_{ik}, \beta, \beta)}{\sum_{z_{mis,ik}} p(x_{ik}, \delta_{ik}|z_{obs,ik}, x_{ik}, \delta_{ik}, \beta, \beta)} p(z_{obs,ik}|x_{ik}, \beta, \beta)
\]

\[
= \frac{[\lambda_0(x_{ik}) \exp(\beta^T z_{ik})]^i_{ik} \exp\{-e^{\beta^T z_{ik}} \Lambda_0(x_{ik})\} p(z_{ik}|\alpha)}{\sum_{z_{mis,ik}} [\lambda_0(x_{ik}) \exp(\beta^T z_{ik})]^i_{ik} \exp\{-e^{\beta^T z_{ik}} \Lambda_0(x_{ik})\} p(z_{ik}|\alpha)}
\]  (18)

Since \( p(x_{ik}, \delta_{ik}|z_{ik}, \lambda, \beta) \) is specified in (3), we must correctly specify and estimate \( p(z_{ik}|\alpha) \) so that (18) is correctly specified. Thus, we need to add an additional set of estimating equations to (11) in order to estimate \( \alpha \).
Now, we discuss estimation of $\alpha$. Even though observations in a cluster can be similar, implying that $Z_{ik}$ and $Z_{ik}$ can be correlated, using the results of Liang and Zeger (1986), with no missing elements of $Z_{ik}$, one can obtain consistent estimates of $\alpha$ by naively assuming independence of the $Z_{ik}$’s within a cluster. Thus, for a given specification for $p(z_{ik}|\alpha)$ and with no missing data, one can consistently estimate $\alpha$ using the maximum likelihood estimating equations naively assuming independence of the $Z_{ik}$’s, $u(\hat{\alpha}) = 0$, where

$$u(\alpha) = \sum_{i=1}^{N} \sum_{k=1}^{n_i} \frac{\partial \log p(z_{ik}|\alpha)}{\partial \alpha}. \tag{19}$$

Again, using method of moments theory, $\hat{\alpha}$ is asymptotically normal and consistent since the score vector $u_{ik}(\alpha)$ has mean 0, and we are solving $u(\hat{\alpha}) = 0$ for $\hat{\alpha}$. The estimating equation (19) alleviates the need to specify a possibly complicated joint distribution for the $Z_{ik}$’s within a cluster. Unfortunately, with missing covariate data, (19) cannot be used. Instead, we propose use of the estimating equations $s_3(\hat{\beta}, \hat{\lambda}, \hat{\alpha}) = 0$, where

$$s_3(\beta, \lambda, \alpha) = \sum_{i=1}^{N} \sum_{k=1}^{n_i} E[u_{ik}(\alpha)|z_{obs,ik}, x_{ik}, \delta_{ik}] = \sum_{i=1}^{N} \sum_{k=1}^{n_i} E \left[ \frac{\partial \log p(z_{ik}|\alpha)}{\partial \alpha} \big| z_{obs,ik}, x_{ik}, \delta_{ik} \right]. \tag{20}$$

Since $E\{E[u_{ik}(\alpha)|z_{obs,ik}, x_{ik}, \delta_{ik}]\} = E[u_{ik}(\alpha)] = 0$, we can again use method of moment theory to show that $\hat{\alpha}$ will be consistent.

Thus, our proposed estimating equations for $\theta = (\beta, \lambda, \alpha)$ are of the form $s(\hat{\beta}, \hat{\lambda}, \hat{\alpha}) = 0$, where

$$s(\theta) = \begin{bmatrix} s_1(\beta, \lambda, \alpha) \\ s_2(\beta, \lambda, \alpha) \\ s_3(\beta, \lambda, \alpha) \end{bmatrix} = \frac{\partial \log p(z_{ik}|\alpha)}{\partial \alpha} \bigg|_{z_{obs,ik}, x_{ik}, \delta_{ik}} \begin{bmatrix} \int_{0}^{\infty} \{ E[Z_{ik}|z_{obs,ik}, x_{ik}, \delta_{ik}] - \bar{Z}_E(s, \beta) \} dN_{ik}(s) \\ dN_{ik}(t) - \lambda(t)Y_{ik}(t)E[e^{\beta z_{ik}}|z_{obs,ik}, x_{ik}, \delta_{ik}] \\ E \left[ \frac{\partial \log p(z_{ik}|\alpha)}{\partial \alpha} \big| z_{obs,ik}, x_{ik}, \delta_{ik} \right] \end{bmatrix}. \tag{21}$$

To make the estimating equations in (21) clearer, for simplicity here, suppose the components of $z_{ik} = [z_{ik1}, ..., z_{ikp}]'$ are discrete. Then, the conditional expectations be-
come sums over conditional probabilities, i.e.,

\[
s(\theta) = \sum_{i=1}^{n_1} \sum_{k=1}^{n_i} \sum_{z_{mis,ik}} p_{ik,z_{mis,ik}} \left[ \int_{0}^{\infty} \left\{ [\mathbf{Z}_{obs,ik}, \mathbf{Z}_{mis,ik}] - \bar{Z}_E(s, \beta) \right\} dN_{ik}(s) \right] dN_{ik}(t) - \lambda_0(t) Y_{ik}(t) e^{\beta z_{jk}}
\]

and

\[
\bar{Z}_E(s, \beta) = \sum_{j=1}^{n} \sum_{k=1}^{n_j} \sum_{z_{mis,ik}} p_{ik,z_{mis,ik}} \frac{Y_{jk}(s) \bar{Z}_{jk} e^{\beta z_{jk}}}{\sum_{j=1}^{n} \sum_{k=1}^{n_j} \sum_{z_{mis,ik}} p_{ik,z_{mis,ik}} Y_{jk}(s) e^{\beta z_{jk}}}
\]

where, for the \(k^{th}\) member of cluster \(i\),

\[
p_{ik,z_{mis,ik}} = pr[\mathbf{Z}_{mis,ik} = z_{mis,ik}|\mathbf{Z}_{obs,ik}, x_{ik}, \delta_{ik}, \theta] = p[\mathbf{z}_{mis,ik}(m)|\mathbf{z}_{obs,ik}, x_{ik}, \delta_{ik}, \theta].
\]

is the conditional probability of a particular value of \(\mathbf{Z}_{mis,ik}\), in (18) and the score vector in (22) is summed over all possible values of \(\mathbf{Z}_{mis,ik}\). These conditional probabilities sum to 1 over all possible patterns of \(\mathbf{z}_{mis,ik}\), i.e., \(\sum_{z_{mis,ik}} p_{ik,z_{mis,ik}} = 1\).

## 5 EM-type Algorithm

### 5.1 Discrete covariates

For discrete \(\mathbf{z}_{mis,ik}\), an obvious choice for a model for \(\mathbf{z}_{ik}\) is one joint multinomial log-linear model (Agresti, 1990). One can then reduce the number of nuisance parameters \(\alpha\) by deleting higher order interaction terms from the log-linear model.

One way to solve \(s(\hat{\theta}) = 0\) in (22) is to use an EM-type algorithm. We can consider
the score vector in (22) as a ‘weighted’ score vector, with weights

\[
p_{ik, zmix, ik} = p_{ik, zmix, ik}(\theta)
\]

\[
= \frac{p(x_{ik}, z_{ik}, z_{ik}, \lambda, \beta)}{\sum_{z_{mis, ik}} p(x_{ik}, z_{ik}, z_{ik}, \lambda, \beta)}
\]

\[
= \frac{[\lambda_0(x_{ik}) \exp(\beta' z_{ik})]^{\delta_{ik}} \exp\left\{-e^{\beta' z_{ik} \Lambda_0(x_{ik})}\right\} p(z_{ik}|\alpha)}{\sum_{z_{mis, ik}} [\lambda_0(x_{ik}) \exp(\beta' z_{ik})]^{\delta_{ik}} \exp\left\{-e^{\beta' z_{ik} \Lambda_0(x_{ik})}\right\} p(z_{ik}|\alpha)}
\]

\[
= \frac{\exp(\beta' z_{obs, ik} z_{mis, ik})^{\delta_{ik}} \exp\left\{-e^{\beta' z_{obs, ik} z_{mis, ik} \Lambda_0(x_{ik})}\right\} p(z_{obs, ik} z_{mis, ik}|\alpha)}{\sum_{z_{mis, ik}} \exp(\beta' z_{obs, ik} z_{mis, ik})^{\delta_{ik}} \exp\left\{-e^{\beta' z_{obs, ik} z_{mis, ik} \Lambda_0(x_{ik})}\right\} p(z_{obs, ik} z_{mis, ik}|\alpha)}.
\]

(25)

We note here that, although the baseline hazard \(\lambda_0\) plays a role in \(p(x_{ik}, \delta_{ik} z_{ik}, \lambda, \beta)\), it is not needed to calculate \(p_{ik, zmix, ik}(\theta)\) because \(\lambda_0(x_{ik})\) depends only on event time \(x_{ik}\), and thus factors out of the numerator and denominator in the conditional probability in (25). Because of this factorization, we avoid the difficulty of ensuring that \(\hat{\lambda}_0(t)\) is consistent (say, by using a smoothing method), and instead only need to ensure that the cumulative hazard \(\hat{\Lambda}(t)\) is consistent.

For a given value of \(\theta = (\beta, \lambda, \alpha)\), say \(\theta^{(l)} = (\beta^{(l)}, \lambda^{(l)}, \alpha^{(l)})\), suppose we define the function

\[
s^x(\theta|\theta^{(l)}) =
\]

\[
\sum_{i=1}^n \sum_{k=1}^{n_i} \sum_{z_{mis, ik}} p_{ik, zmis, ik}^{(l)} \left[ \int_0^\infty \left\{ [Z_{obs, ik}, Z_{mis, ik}] - Z_E(s, \beta) \right\} dN_{ik}(s) \right] \frac{dN_{ik}(t) - \lambda_0(t)Y_{ik}(t) e^{\beta' [z_{obs, ik}, z_{mis, ik}]}}{\partial \log p_{z_{obs, ik}, z_{mis, ik}|\alpha}}.
\]

(26)

where \(p_{ik, zmis, ik}^{(l)} = p_{ik, zmis, ik}(\theta^{(l)})\). The EM-type iterative algorithm to solve \(s(\hat{\beta}, \hat{\lambda}, \hat{\alpha}) = 0\) entails the following:

- Obtain an initial estimate \(\theta = \theta^{(1)}\), say, by complete cases. At the \(\ell^{th}\) step, we have \(\theta^{(l)}\).
- Using \(\theta^{(l)}\), calculate \(p_{ik, zmis, ik}^{(l)} = p_{ik, zmis, ik}(\theta^{(l)})\).
• Fixing $p_{ik,z_{mis,ik}}^{(\ell)} = p_{ik,z_{mis,ik}}(\theta^{(\ell)})$, solve $s(\theta^{(\ell+1)}|\theta^{(\ell)}) = 0$ for $\theta^{(\ell+1)}$.

• We iterate until convergence, which gives the solution to $s(\hat{\theta}) = 0$.

Although the regression parameters estimates are consistent when naively assuming the observations in a cluster as independent, when the survival times within a cluster are correlated, the inverse of the information matrix under independence may be an inconsistent estimate of the asymptotic variance. The asymptotic variance of $\hat{\theta}$ can be consistently estimated using a robust ‘sandwich’ estimator (White, 1982),

$$\hat{V}ar(\hat{\theta}) = \left\{ \sum_{i=1}^{N} \sum_{k=1}^{n_i} \hat{s}_{ik}(\hat{\theta}) \right\}^{-1} \left\{ \sum_{i=1}^{N} \left[ \sum_{k=1}^{n_i} s_{ik}(\hat{\theta}) \right] \left[ \sum_{k=1}^{n_i} s_{ik}(\hat{\theta})' \right] \right\} \left\{ \sum_{i=1}^{N} \sum_{k=1}^{n_i} \hat{s}_{ik}(\hat{\theta}) \right\}^{-1} \quad (27)$$

where

$$\hat{s}_{ik}(\hat{\theta}) = \left[ \frac{\partial s_{ik}(\theta)}{\partial \theta} \right]_{\theta = \hat{\theta}}.$$ 

The upper $P \times P$ block of this estimate is consistent for the asymptotic variance of $\hat{\beta}$. Alternatively, one can use a proof similar to (Lipsitz, one can use the jackknife to estimate the variance, i.e.,

$$\hat{V}ar(\hat{\theta}) = \sum_{i=1}^{n} (\hat{\theta}_{-i} - \hat{\theta})(\hat{\theta}_{-i} - \hat{\theta})',$$

where $\hat{\theta}_{-i}$ is the estimate of $\theta$ obtained after dropping the $i^{th}$ cluster.

5.2 Continuous covariates

For continuous or mixed $z_{mis,ik}$, one need not pose a multivariate normal distribution or combination of multivariate normal and log-linear model, but can pose a set of any continuous or discrete univariate distributions, as described in Ibrahim, Chen and Lipsitz (1999). When the missing covariates are continuous, one can adapt the Monte Carlo EM algorithm of Wei and Tanner (1990) or the stochastic approximation Monte Carlo EM
algorithm of Gu and Lin (1998). With $z_{\text{mis},ik}$ continuous, the score vector in (21) now becomes

$$s(\theta) = \sum_{i=1}^{n_i} \sum_{k=1}^{n_i} \int_{z_{\text{mis},ik}} p_{ik,z_{\text{mis},ik}} \left[ \int_{0}^{\infty} \left\{ \left[ Z_{\text{obs},ik}, Z_{\text{mis},ik} \right] - \bar{Z}_E(s, \beta) \right\} dN_{ik}(s) \right] \frac{dN_{ik}(t) - \lambda_0(t)Y_{ik}(t)e^{\beta'z_{\text{obs},ik,z_{\text{mis},ik}}}}{\partial \log p(z_{\text{obs},ik,z_{\text{mis},ik}}|\alpha)} \partial \alpha \right], \quad (28)$$

where $p_{ik,z_{\text{mis},ik}}$ is now the conditional density

$$p_{ik,z_{\text{mis},ik}} = \exp(\beta'z_{\text{obs},ik,z_{\text{mis},ik}}) \delta_{ik} \exp \left\{ -e^{\beta'z_{\text{obs},ik,z_{\text{mis},ik}}} \lambda_0(x_{ik}) \right\} p(z_{\text{obs},ik,z_{\text{mis},ik}}|\alpha). \quad (29)$$

To obtain the solution $s(\hat{\theta}) = 0$, we draw $M$ values of $z_{\text{mis},ik}$ from $p_{ik,z_{\text{mis},ik}}^{(\ell)}$, which is (29) evaluated at $\theta^{(\ell)}$. However, because (29) often has no closed form, we can draw these values of $z_{\text{mis},ik}$ using the Gibbs Sampler (Gilks and Wild, 1992). We estimate $s(\theta|\theta^{(\ell)})$ with

$$s^*(\theta|\theta^{(\ell)}) = \sum_{i=1}^{n_i} \sum_{k=1}^{n_i} \frac{1}{M} \left[ \int_{0}^{\infty} \left\{ \left[ Z_{\text{obs},ik}, Z_{\text{mis},ik}^{m(\ell)} \right] - \bar{Z}_E(s, \beta)^{m(\ell)} \right\} dN_{ik}(s) \right] \frac{dN_{ik}(t) - \lambda_0(t)Y_{ik}(t)e^{\beta'z_{\text{obs},ik,z_{\text{mis},ik}^{m(\ell)}}}}{\partial \log p(z_{\text{obs},ik,z_{\text{mis},ik}}^{m(\ell)}|\alpha)} \partial \alpha \right], \quad (30)$$

where $z_{\text{mis},ik}^{m(\ell)}$ is $m^{th}$ draw from $p_{ik,z_{\text{mis},ik}}^{(\ell)}$. In iterations of Monte-Carlo EM algorithm, instead of solving $s(\theta|\theta^{(\ell)}) = 0$ for $\theta^{(\ell+1)}$, we solve its approximation $s^*(\theta|\theta^{(\ell)}) = 0$. Then, the EM-type algorithm proceeds just as with discrete $z_{\text{mis},ik}$.

### 6 Example

Although there are some difficulties in determining the precise number of deaths, each year approximately 2500 persons in the United States die of liver cancer. The Eastern Cooperative Oncology Group (ECOG) has undertaken a series of phase II clinical trials to
evaluate new treatments in patients with primary liver cancer. To illustrate our proposed methods, we consider data from two such clinical trials, EST 2282 (Falkson et al. 1990) and EST 1286 (Falkson et al., 1994). We analyze a sample of 182 patients from 31 hospitals (clusters). In similar studies, it has often occurred that survival time from subjects within the same hospital tend to be correlated. Then, we can think of a hospital as cluster $i$, with $n = 31$; within the $i^{th}$ institution, we have $n_i$ patients (with 182 patients, we have an average of 5.9 patients per institution). The observed response for the $k^{th}$ patient in institution $i$ is the minimum of the censoring and survival time.

We are primarily interested in how survival, time from entry on the study until death, differs with respect to five baseline characteristics. These five baseline characteristics are age (less than 60 or greater than or equal to 60); associated jaundice (yes, no); associated hepatitis (yes, no); and two biochemical markers Alpha fetoprotein and Anti Hepatitis B antigen, each classified as normal or abnormal. Older patients, patients with abnormal biochemical markers, jaundice, and/or hepatitis are all expected to have shorter survival. Each of the five covariates are coded 0 or 1, with 1 assigned to the category expected to lead to shorter survival. Unfortunately, as shown in Table 2, 25.3% of the patients have at least one covariate missing. The biochemical markers, which require blood to be drawn and possibly complicated laboratory tests, have the highest proportion missing.

Assuming a Cox proportional hazards model with non-informative censoring, we have

$$p(x_{ik}, \delta_{ik}|z_{ik}, \beta) = [\lambda_0(x_{ik})\lambda_{ik}]^{\delta_{ik}} \exp\{-\lambda_{ik}\Lambda_0(x_{ik})\}$$

(31)

where

$$\lambda_{ik} = \exp(\beta'z_{ik}) .$$

Here, we are interested in a main effects model, where $x_{ik} = (x_{ik1}, ..., x_{ik5})'$ contains the five dichotomous covariates. Since all covariates are discrete, we model $p(x_{ik}|\alpha)$ as a
saturated multinomial with $2^5$ levels.

Before we fit the Cox model, we would like to briefly see if the missing data mechanism for these data fits the assumptions given in (5) and (6). To test if (5) holds versus an alternative of non-ignorable missingness (i.e., missingness depends on $x_{\text{mis},ik}$), we would need to fit a complicated joint likelihood of the survival time, missing data indicators, and covariate distributions; however, the main advantage of the method described here is to avoid having to specify and maximize such a likelihood. Further, such non-ignorable models are often non-identifiable, and lead to inestimable parameters (Baker and Laird, 1988). Thus, here we are interested in testing if the missing data are missing completely at random (i.e., missingness does not depend on $x_{ik}, \delta_{ik},$ or $z_{\text{obs},ik}$) versus missing at random.

To do this in the most rigorous sense, we would still need to fit a complicated joint likelihood of the survival times, the missing data indicators, and covariate distributions, which is again beyond the scope of this paper. Instead, to test if the covariates are missing completely at random, we define a simpler version of (6) which can be written as follows. Let the indicator $R_{ik} = \prod_{p=1}^{P} R_{ikp}$ equal 1 when the $k^{th}$ member of cluster $i$ has no missing covariate data and 0 otherwise. We then form a logistic regression model for the probability that $R_{ik} = 1$ as a function of the completely observed variables, jaundice, follow-up time, and censoring time (any of the other covariates would have at least one missing value in this model). Further, since only one subject has a censored outcome (the last patient in Table 1), we could not fit a logistic regression for $R_{ik}$ with censoring as a covariate. Thus, we model the probability of being a complete case ($R_{ik} = 1$) as a function of the completely observed data,

$$\logit[\text{pr}(R_{ik} = 1|\text{jaun}_{ik}, s_{ik}, \alpha)] = \alpha_0 + \alpha_1 s_{ik} + \alpha_2 \text{jaun}_{ik} + \alpha_1 s_{ik} \text{jaun}_{ik}, \quad (32)$$

where $s_{ik}$ equals 1 if the subject lived longer than 12 weeks, and equals 0 if the subject died within the first 12 weeks. We actually fit this model using generalized estimating
equations (Liang and Zeger, 1986), assuming \( R_{ik} \)'s in the same cluster had an exchangeable correlation, i.e.,

\[
\rho = \text{Corr}(R_{ij}, R_{ik}|\text{jaun}_{ij}, s_{ij}, \text{jaun}_{ik}, s_{ik}).
\]

The estimate of \( \alpha \) and \( \rho \) are given in Table 3, using the robust estimate of variance of \( \hat{\alpha} \) and \( \hat{\rho} \) given in Prentice (1988).

Since the main effects of jaundice and dichotomized survival, and their interaction are significant in Table 3, this suggests that the missing data are not missing completely at random. In particular, the probability of being a complete case depends on both the jaundice status and survival. A patient without jaundice and with a short survival has the smallest estimated probability of being a complete case, equal to 0.46. A patient with both jaundice and a longer survival has the next smallest estimated probability of being a complete case, equal to 0.76. A patient with jaundice and with a short survival has estimated probability of being a complete case equal to 0.79. Finally, a patient without jaundice and with a longer survival has the highest estimated probability of being a complete case, equal to 0.81. On average, patients with jaundice (compared to no jaundice) are more likely to be complete cases, possibly due to the fact that doctors of patients with jaundice may want to see all of the laboratory tests and biochemical markers at baseline. Also, on average, patients with survival greater than 12 weeks (compared to those less than or equal to 12 weeks) are more likely to be complete cases, possibly due to the fact that the data managers have more time to query the laboratories for the values of the baseline biochemical markers for a patient who lives longer; it may not be worth the trouble to update the file of a patient who has already passed away. Further, doctors may order follow-up laboratory tests for a patient who lives longer (or have a longer censoring time), and thus a data manager queried the laboratory for the baseline values in order to look at the change from baseline. Since the intracluster correlation in Table 3 is non-
significant, this suggests that (6) and not (5), holds, i.e., that $r_{ik}$ given all data for cluster $i$ depends only on the observed data from patient $k$ in cluster $i$.

Table 4 gives the estimates of $\beta$. The jackknife was used to estimate the variances of the complete case (CC) and our proposed (EM) estimate, under both the assumption of independence and clustering. In particular, under the assumption of independence, each of the 182 patients were dropped out one at a time, with the jackknife variance calculated from their variability; under the assumption of clustering, each of the 31 clusters were dropped out one at a time. The CC and EM parameter estimates in Table 4 are very similar, except for the Jaundice effect, which is more than twice the magnitude (-.295 for EM versus -.133 for CC). This is not surprising in that, from Table 3, the probability of being a complete case depends on both the censoring time and jaundice, so that we would expect the complete case estimate of jaundice to be biased. Using the $p-$values under clustering, we see that the Jaundice effect using EM is significant at the 5% level ($p = 0.025$), whereas the CC estimate is not significant ($p = 0.461$). Further, when just considering the EM estimate of the Jaundice effect, we see that the variance estimate under clustering (0.0174) is 40% smaller than the corresponding variance estimate (0.0282) under the naive assumption of independence; because of this, the $p-$value under the naive assumption of independence is not significant at the 5% level ($p = 0.078$). Thus, one can get different conclusions if one does not take the possible clustering of patients within a hospital into account. Further, although the $p-$value is greater than 5% in both cases, for the age effect, the variance estimate under clustering (0.0071) is only 25% of the value of the corresponding variance estimate (0.0269) under the naive assumption of independence. As we see from the variance estimates, it is not always the case that the variance estimates under independence will be smaller than the robust variance estimate. Overall, this example shows that using the usual variance estimate under independence...
can give misleading results. This demonstrates how an analysis that does not use the robust variance estimate for clustering can give misleading and conflicting results.

To use the method here, one can use the usual EM-algorithm for independent observations. To find our proposed estimate \( \hat{\beta} \), we iterate between the E and M steps until \( \beta^{(m+1)} = \beta^{(m)} = \hat{\beta} \). We note here that the convergence criterion used for the EM-type algorithm was that the distance between the \( m^{th} \) iteration and the \((m + 1)^{th}\) iteration in each parameter was less than \(10^{-5}\). The number of iterations required for convergence was 12, with the complete case estimates used as starting values. We programmed the algorithm in SAS, and our SAS macro took 3 seconds in real time to calculate the EM-type estimate on a Pentium 4, 2.4 GHZ PC with 512 MB of RAM.

7 Discussion

In this paper, we have shown that if one naively treats observations within a cluster as independent, that one can still use the EM-type algorithm of Herring and Ibrahim (2001) to obtain consistent estimates of the relative risk parameters, but one must use a robust variance estimator. Although the estimates are consistent, they could be inefficient if there is high correlation between observations in a cluster. A useful extension would then be to develop estimating equations that somehow use the correlation between the clustered censored survival times. Unfortunately, there is not even a good estimating equations approach for estimating the intra-correlation coefficient with clustered censored survival data with no missing covariates (Segal et al., 1997). Adding missing covariates on top of the censoring make this a daunting task. Alternatively, the most logical approach is a full likelihood approach, in which we specify a joint distribution for the clustered survival times given the covariates, and a marginal model for the covariates. The numerical optimization
for such models without missing covariates is quite complicated. Further, mis-specification of the likelihood could lead to biased estimates of the relative risk parameters, whereas our approach only requires correct specification of the marginal distribution. Thus, our approach is a relatively simple and useful method to use with clustered survival data with missing covariates.

ACKNOWLEDGEMENTS

The authors are grateful for the support provided by the following grants from the United States’ Institutes of Health: CA 79060, HL 69800, AHRQ 10871, HL52329, HL61769, CA 70101, and CA 74015.

REFERENCES


<table>
<thead>
<tr>
<th>Cluster</th>
<th>Survival (weeks)</th>
<th>Age</th>
<th>Jaundice</th>
<th>Hepatitis</th>
<th>Abnormal Anti-Hep</th>
<th>Abnormal Alpha feto</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.000 ≥ 60</td>
<td>NO</td>
<td>-</td>
<td>YES</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6.000 ≥ 60</td>
<td>YES</td>
<td>NO</td>
<td>-</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.143 ≥ 60</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14.571 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>187.857 ≥ 60</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.857 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11.000 &lt; 60</td>
<td>YES</td>
<td>NO</td>
<td>-</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>21.429 &lt; 60</td>
<td>NO</td>
<td>YES</td>
<td>-</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>60.857 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5.571 &lt; 60</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9.429 &lt; 60</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12.000 &lt; 60</td>
<td>NO</td>
<td>NO</td>
<td>-</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>14.571 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>28.286 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10.143 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>13.714 ≥ 60</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>15.000 &lt; 60</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.857 &lt; 60</td>
<td>YES</td>
<td>NO</td>
<td>-</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11.857 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16.714 &lt; 60</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.857 &lt; 60</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8.714 &lt; 60</td>
<td>NO</td>
<td>YES</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9.000 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>25.143 &lt; 60</td>
<td>NO</td>
<td>NO</td>
<td>-</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>25.143 &lt; 60</td>
<td>NO</td>
<td>NO</td>
<td>-</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>13.714 ≥ 60</td>
<td>YES</td>
<td>NO</td>
<td>-</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>27.571 ≥ 60</td>
<td>YES</td>
<td>-</td>
<td>YES</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>30.714</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>99.429 &lt; 60</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>124.714 &lt; 60</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3.857 ≥ 60</td>
<td>NO</td>
<td>NO</td>
<td>-</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>449.143* &lt; 60</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>

*=censored
- = missing
<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>PERCENT MISSING</th>
<th>NUMBER MISSING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>Anti-Hep</td>
<td>19.2</td>
<td>35</td>
</tr>
<tr>
<td>Alpha feto</td>
<td>7.1</td>
<td>13</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>1.6</td>
<td>3</td>
</tr>
<tr>
<td>Age</td>
<td>1.1</td>
<td>2</td>
</tr>
<tr>
<td>Jaundice</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>OVERALL</strong></td>
<td><strong>25.3</strong></td>
<td><strong>46</strong></td>
</tr>
</tbody>
</table>
Table 3. Estimates for the missing data model (for probability of being observed, $R_{ik} = 1$)

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>ESTIMATE</th>
<th>ERROR</th>
<th>Z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>0.163</td>
<td>0.428</td>
<td>-0.38</td>
<td>0.704</td>
</tr>
<tr>
<td>JAUN</td>
<td>1.468</td>
<td>0.573</td>
<td>2.56</td>
<td>0.010</td>
</tr>
<tr>
<td>$(T_{ik} &gt; 12)$</td>
<td>1.638</td>
<td>0.717</td>
<td>2.28</td>
<td>0.022</td>
</tr>
<tr>
<td>JAUN $(T_{ik} &gt; 12)$</td>
<td>-1.823</td>
<td>0.903</td>
<td>-2.02</td>
<td>0.044</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.009</td>
<td>0.029</td>
<td>0.29</td>
<td>0.769</td>
</tr>
</tbody>
</table>
Table 4. EM Estimates for $\beta$ for Cox regression model$^a$

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>METHOD</th>
<th>ESTIMATE</th>
<th>STANDARD ERROR</th>
<th>Z–value</th>
<th>p–value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha feto</td>
<td>EM</td>
<td>0.291</td>
<td>0.180 (0.189)</td>
<td>1.61</td>
<td>0.106 (0.123)</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>0.296</td>
<td>0.165 (0.199)</td>
<td>1.79</td>
<td>0.073 (0.137)</td>
</tr>
<tr>
<td>Anti–Hep</td>
<td>EM</td>
<td>-0.006</td>
<td>0.372 (0.285)</td>
<td>-0.02</td>
<td>0.987 (0.983)</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>-0.001</td>
<td>0.393 (0.320)</td>
<td>0.01</td>
<td>0.998 (0.997)</td>
</tr>
<tr>
<td>Age</td>
<td>EM</td>
<td>0.086</td>
<td>0.084 (0.164)</td>
<td>1.02</td>
<td>0.310 (0.602)</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>0.084</td>
<td>0.096 (0.203)</td>
<td>0.88</td>
<td>0.380 (0.680)</td>
</tr>
<tr>
<td>Jaundice</td>
<td>EM</td>
<td>-0.295</td>
<td>0.132 (0.168)</td>
<td>-2.24</td>
<td>0.025 (0.078)</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>-0.133</td>
<td>0.180 (0.185)</td>
<td>-0.74</td>
<td>0.461 (0.471)</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>EM</td>
<td>0.325</td>
<td>0.383 (0.366)</td>
<td>0.85</td>
<td>0.396 (0.375)</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>0.255</td>
<td>0.382 (0.416)</td>
<td>0.67</td>
<td>0.505 (0.541)</td>
</tr>
</tbody>
</table>

$^a$ statistics under independence assumption in parentheses