Maximum likelihood estimation in random effects
cure rate models with nonignorable missing
covariates

AMY H. HERRING*
Department of Biostatistics, The University of North Carolina at Chapel Hill, Campus Box 7420,
Chapel Hill, NC 27599, USA
aherring@bios.unc.edu

JOSEPH G. IBRAHIM
Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute,
655 Huntington Avenue, Boston MA 02115, USA

SUMMARY

We introduce a method of parameter estimation for a random effects cure rate model. We also propose
a methodology that allows us to account for nonignorable missing covariates in this class of models. The
proposed method corrects for possible bias introduced by complete case analysis when missing data are
not missing completely at random and is motivated by data from a pair of melanoma studies conducted
by the Eastern Cooperative Oncology Group in which clustering by cohort or time of study entry was
suspected. In addition, these models allow estimation of cure rates, which is desirable when we do not
wish to assume that all subjects remain at risk of death or relapse from disease after sufficient follow-up.
We develop an EM algorithm for the model and provide an efficient Gibbs sampling scheme for carrying
out the E-step of the algorithm.

Keywords: Cure rate model; Gibbs sampling; Missing covariates; Monte Carlo EM algorithm; Nonignorable missing
data; Random effects; Survival analysis.

1. INTRODUCTION

Because of improvements in the treatment of cancer, cure rate models have become increasingly
popular in the analysis of data from cancer clinical trials. For certain cancers, including breast cancer,
leukemia, melanoma, and prostate cancer, a substantial proportion of patients may now be cured by
treatment. Traditional methods of survival analysis, including the Cox (1972) regression model, assume
that no patients are ‘cured’ but that all remain at risk of death or relapse. These models are concerned
with survival only and do not address the possibility of disease cure. However, estimation of a treatment-
specific cure rate provides valuable information that is not only of use to the investigator but is of primary
interest to the patient at the time of diagnosis.

Analysis of two melanoma clinical trials heightened our interest in cure rate models. Melanoma
incidence rates are increasing faster than those for any other solid tumor. While education and greater
awareness have led to a cure rate of up to 85% for early stage disease, later stage disease can be lethal, with relapse and mortality rates of 60–75% for high-risk patients. The Eastern Cooperative Oncology Group (ECOG) carried out clinical trial E1684 to assess the effect of high-dose interferon as post-operative chemotherapy for patients with high-risk melanoma. This trial showed a substantial treatment benefit due to interferon and led to US FDA approval of this regimen, but because this benefit was much larger than anticipated and was accompanied by substantial side effects, ECOG decided to carry out a confirmatory trial, E1690. No significant benefit of treatment was found in the confirmatory trial.

These clinical trials were initially analyzed using standard survival methods that did not allow estimation of cure rates. While the probability of cure for high-risk melanoma is much lower than the cure rates seen for early stage disease, estimation of these fractions was of great interest to the investigators. In addition, investigators suspected that advances in treatment and care (e.g. improved surgical techniques) on both treatment arms were prolonging survival as the studies progressed. We introduce a random effect in the cure rate model to account for a possible cohort or time of entry effect. Further complicating the analysis is the fact that several tumor characteristics related to survival were missing for a substantial number of patients. The goals of our analysis were to account properly for the missing data (and to attempt to assess whether the missing data were not MCAR, implying that a complete case analysis could be biased), to estimate cure rates, and to incorporate a random effect for cohort. More details and our complete analysis are given in Section 4.

Consider the survival functions given in Figure 1. This figure shows estimated population survival curves by treatment group for the Cox (1972) proportional hazards model and a cure rate model fit to our melanoma data. We see that although the two models provide similar survival estimates early in the study, the Cox model does a poor job estimating the tails of the survival curve due to the lack of events later in the study. The cure rate model provides smoother survival estimates in the tail, as these estimates are not as sensitive to the lack of events at later times. In addition, the cure rate model enables us to estimate the survival curves for the group of patients not ‘cured’ by the treatment, a quantity not available from the Cox model. These survival curves would apply specifically to those patients still at risk of failure, while the population survival curves presented in Figure 1 apply to ‘cured’ and ‘non-cured’ patients alike. This issue is discussed in more detail in the analysis of the melanoma data given in Section 4.
As previously mentioned, the positive probability of cure is not the only factor complicating the analysis. Event times are frequently not independent, as events may be correlated because certain subgroups share common traits, such as cage or litter effects in animal studies, cohort or institutional effects in clinical trials, or measurements of repeated events within subjects. By introducing random effects for these subgroups, we are able to obtain proper variance estimates by accounting for the homogeneity of subjects within subgroups.

Even well-planned clinical trials and observational studies frequently have large fractions of missing covariate data. The most popular method for dealing with missing covariates is a complete case analysis, which analyzes only those subjects with complete data and discards those with missing covariate values. This method of analysis is the default method despite the knowledge that this analysis makes strong (and generally unverifiable) assumptions about the missing data. Large fractions of missing data lead to inefficient parameter estimates and may introduce serious biases into a complete case analysis, biases that may be especially severe when the missing covariates are nonignorable. Several authors have addressed the problem of missing data in the Cox regression model (Lin and Ying, 1993; Paik, 1997; Paik and Tsai, 1997; Lipsitz and Ibrahim, 1998; Chen and Little, 1999; Herring and Ibrahim, 2001), but they do not consider survival data with the cure rate structure and do not allow missing covariates to be nonignorable. Chen and Ibrahim (2001) recently addressed missing at random covariates in cure rate models but do not allow for nonignorable missing data or random effects.

We propose a method for estimating the parameters of a semi-parametric survival model with a cure fraction. Our model incorporates random effects and accounts for nonignorable missing covariates. Like the models of Chen et al. (1999) and Chen and Ibrahim (2001), the hazard for the entire population has the proportional hazards structure, even when covariates are included. The random effects are introduced as part of the linear predictor, a formulation similar to the one commonly used in generalized linear mixed models, and the random effects may come from a large family of distributions.

The rest of this paper is organized as follows. In Section 2 we propose a Monte Carlo EM algorithm for obtaining parameter estimates when all covariates are observed. Then we extend the methodology to account for nonignorable missing covariates, and we obtain standard errors using a multiple imputation method (Goetghebeur and Ryan, 2000) in Section 3. We present our analysis of the melanoma data in Section 4 along with a detailed sensitivity analysis of our modeling assumptions and conclude in Section 5 with a discussion of the analysis and issues worthy of further consideration.

2. COMPLETE DATA RANDOM EFFECTS CURE RATE MODEL

We adapt the cure rate model proposed by Chen et al. (1999) to accommodate correlated event times and nonignorable missing covariates. This model is related to the standard cure rate model, addressed in the literature by many authors including Berkson and Gage (1952); Ewell and Ibrahim (1997); Farewell (1982, 1986); Goldman (1984); Gray and Tsiatis (1989); Greenhouse and Wolfe (1984); Halpern and Brown (1987a,b); Kuk and Chen (1992); Laska and Meisner (1992); Maller and Zhou (1996); Sposto et al. (1992); Stangl and Greenhouse (1998); Taylor (1995), and Yamaguchi (1992). Our formulation of the cure rate model, as opposed to the standard cure rate model, retains a proportional hazards structure for the entire population in the presence of covariates, giving the model an attractive interpretation and biological motivation described below.

We extend the model of Chen et al. (1999) to clustered observations. To our knowledge, no other authors have addressed random effects cure rate models with missing data. Let \( N_{hi} \) denote the number of metastasis-competent tumor cells after the initial cancer treatment for the \( i \)th patient in the \( h \)th group, where \( N_{hi} \sim \text{Poisson}(\theta_{hi}) \). These unobserved cell counts are treated as latent variables in the model. Let \( Z_{hij} \) denote the random promotion time for the \( j \)th carcinogenic cell, \( j = 1, \ldots, N_{hi} \), and assume that for
a given group $h$ and number of cells $N_{hi}$, the variables $Z_{hij}$ are i.i.d. with common distribution function $F(t)$. The time to cancer relapse is the minimum promotion time over all the metastasis-competent cells and may be represented by the random variable $T_{hi} = \min \{Z_{hij}, 1 \leq j \leq N_{hi}\}$. Let $y_h = (y_{h1}, \ldots, y_{hn_h})$ be the observed event times for the $h$th cluster, where $n_h$ is the number of subjects in the $h$th cluster, $h = 1, \ldots, H$, and denote the corresponding failure or censoring indicators by $\nu_h = (\nu_{h1}, \ldots, \nu_{hn_h})$. Let $x_h = (x_{h1}, \ldots, x_{hn_h})'$ be an $n_h \times p$ matrix of fixed covariates for the subjects in the $h$th cluster and let $w_h = (w_{h1}, \ldots, w_{hn_h})'$ be an $n_h \times q$ matrix of covariates that are multiplied by the $q \times 1$ vector of random effects $b_h$. For random intercept models, $w_h$ will simply indicate cluster membership. For convenience we let the i.i.d. random effects follow a multivariate normal distribution; that is, $b_h \sim N_q(0, \Sigma)$. Although it is not necessary to choose a multivariate normal distribution for the random effects, the Gibbs sampling procedure is greatly simplified if we restrict our choice to distributions that are log-concave.

The population survival function is given by $S_p(t) = P(\text{cancer free at time } t) = \exp(-\theta_{hi}F(t))$, where $\theta_{hi} = \exp(x_{hi}'\beta + w_{hi}'b_h)$. Note that $S_p(t)$ is not a proper survival function as $S_p(\infty) = \exp(-\theta_{hi}) \neq 0$. The quantity $S_p(\infty) = \exp(-\theta_{hi}) = \exp(-\exp(x_{hi}'\beta + w_{hi}'b_h))$ is interpreted as the cure fraction and represents the proportion of patients who are ‘cured’. As $\theta_{hi} \to \infty$, the cure fraction tends to 0, while as $\theta_{hi} \to 0$, the cure fraction tends to 1. We consider a piecewise exponential model for $F(t)$, partitioning the time axis as follows. Choosing $s_K$ greater than the maximum event time, we construct the $K$ intervals $[0, s_1], (s_1, s_2], \ldots, (s_{K-1}, s_K]$. Assuming a constant hazard $\lambda_k$ in the $k$th interval, we write $F(t)$ as

$$F^*(t) = 1 - \exp \left\{ -\lambda_k(t - s_{k-1}) - \sum_{g=1}^{k-1} \lambda_g (s_g - s_{g-1}) \right\}, \quad (2.1)$$

and $S^*(t) = 1 - F^*(t)$. When $K = 1$, $F^*(t)$ is simply the cdf of the exponential distribution. For smaller $K$, we get a more parametric model for $F^*(t)$; for larger $K$, we get a more nonparametric shape. We recommend evaluating several values of $K$ to ensure the model is not too sensitive to the specific value of $K$ that is chosen.

Let $\gamma = (\beta, \lambda, \Sigma)$. The complete data are $D_{\text{comp}} = (D_{hi,\text{comp}}, h = 1, \ldots, H, i = 1, \ldots, n_h)$, where $D_{hi,\text{comp}} = (y_{hi}, \nu_{hi}, N_{hi}, x_{hi}, w_{hi}, b_h)$. The complete data log-likelihood is written

$$l(\gamma \mid D_{\text{comp}}) = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \nu_{hi} \log(N_{hi} f^*(y_{hi} \mid \lambda)) + (N_{hi} - \nu_{hi}) \log(S^*(y_{hi} \mid \lambda)) + \sum_{h=1}^{H} \sum_{i=1}^{n_h} N_{hi}(x_{hi}'\beta + w_{hi}'b_h) - \log(N_{hi}) - \exp(x_{hi}'\beta + w_{hi}'b_h)
\quad + \sum_{h=1}^{H} -q \frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2} (b_h'\Sigma^{-1}b_h)
\quad = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \nu_{hi} \log(N_{hi}) + \nu_{hi} \log(\lambda_{hi}) + N_{hi} \log(S^*(y_{hi} \mid \lambda)) + \sum_{h=1}^{H} \sum_{i=1}^{n_h} N_{hi}(x_{hi}'\beta + w_{hi}'b_h) - \log(N_{hi}) - \exp(x_{hi}'\beta + w_{hi}'b_h)
\quad + \sum_{h=1}^{H} -q \frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2} (b_h'\Sigma^{-1}b_h). \quad (2.2)$$

We obtain $f^*$ and $S^*$ using (2.1). More detail about $f^*$ and $S^*$ is provided in the Appendix.
To obtain parameter estimates with no missing covariate data, we implement an EM algorithm and take the conditional expectation, given the observed data, of the log-likelihood in (2.2) with respect to the unobserved random effects and the latent cell counts $N_{hi}$. This involves sampling the random effects at each EM iteration and then estimating the other parameters while treating the sampled random effects as fixed. Let $\gamma^{(l)} = (\beta^{(l)}, \lambda^{(l)}, \Sigma^{(l)})$ at the $l$th iteration of the EM algorithm. The E-step at the $(l+1)$th EM iteration is given by

$$E[l(\gamma' | D_{\text{comp}}) | x_{hi}, \gamma^{(l)}] = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \int_{b_h, N_{hi}} p(b_h, N_{hi} | x_{hi}, \gamma^{(l)}) [v_{hi} \log(N_{hi})]
+ v_{hi} \log(\lambda_{khi}) + N_{hi} \log(5^*(y_{hi} | \lambda^{(l)})) \, db_h \, dN_{hi}$$

$$+ \sum_{h=1}^{H} \sum_{i=1}^{n_h} \int_{b_h, N_{hi}} p(b_h, N_{hi} | x_{hi}, \gamma^{(l)}) [N_{hi}(x'_{hi} \beta + w'_{hi} b_h) - \log(N_{hi}!)]$$

$$- \exp(x'_{hi} \beta + w'_{hi} b_h) \, db_h \, dN_{hi} + \sum_{h=1}^{H} \int_{b_h, N_{hi}} p(b_h, N_{hi} | x_{hi}, \gamma^{(l)})$$

$$\times \left\{ - \frac{q}{2} \log(2\pi) - \frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2} (b'_h \Sigma^{-1} b_h) \right\} \, db_h \, dN_{hi}, \quad (2.3)$$

where $p(b_h, N_{hi} | x_{hi}, \gamma^{(l)})$ is the joint density of the random effects and latent count variables conditional on the data. We outline the evaluation of the E-step using Monte Carlo EM in the Appendix. When covariates are missing, estimation becomes more complicated, as we see in the next section. Variance estimation for both the complete data and missing data cases is outlined in the following section as well.

### 3. Random Effects Cure Rate Model with Missing Data

To obtain valid parameter estimates when covariates are missing, it is necessary to make additional assumptions about the data regardless of whether one plans to conduct a complete case analysis or a more sophisticated analysis. A complete case analysis, except in very special cases, generally requires that data are missing completely at random (MCAR). If MCAR is not the correct missing data mechanism, then a complete case analysis may be biased. One way to obtain valid inference with non-MCAR missing data mechanisms is to specify correctly the distribution of covariates that are not completely observed and possibly also the non-MCAR missing data mechanism. We stress that although the investigator is never absolutely certain that these distributions have been specified correctly except in very special situations, it is also true that investigators can rarely be absolutely certain that data are truly MCAR. This underscores the necessity of presenting a detailed sensitivity analysis along with any statistical results. When the data are missing at random (MAR), we specify the distribution of the missing covariates themselves, and when the covariates are nonignorable, we model the missing data mechanism in addition to the covariate distribution. Although these assumptions go beyond those typical in analyses when all covariates are observed, we cannot proceed with only the standard assumptions and pay no price for having missing data. Our additional model assumptions, needed to obtain unbiased estimates when data are not MCAR, are clearly specified, and the sensitivity of parameter estimates to the modeling assumptions may be checked by varying the covariate and missing data models in a sensitivity analysis, such as the one presented in Section 4.
3.1 Specification of the covariate distribution

When some of the covariates have missing values, we will specify a parametric distribution for $x_{\text{mis}}$ and estimate its parameters from the data, viewing these as nuisance parameters not of inferential interest. Because parameter estimation may become too computationally intensive and unstable with many nuisance parameters, we will need to employ strategies to reduce the number of nuisance parameters in the specification of the covariate distribution as suggested by Lipsitz and Ibrahim (1996) and Ibrahim et al. (1999b), who propose modeling the joint distribution of the missing covariates as a product of one-dimensional conditional distributions.

Let $x_{hi}$ be independent and identically distributed random covariate vectors with density $p(x_{hi} | \alpha)$, where $\alpha$ is distinct from $\gamma = (\beta, \lambda, \Sigma)$. Following Lipsitz and Ibrahim (1996), we model the joint distribution of the covariates as a product of one-dimensional conditional distributions. For ease of expression, suppose that $x_{hi} = (x_{hi1}, \ldots, x_{hir})$, where $(x_{hi1}, \ldots, x_{hir})$ are missing for at least one subject, and $x_{hi, \text{obs}} = (x_{hi1},r+1, \ldots, x_{hir})$ are observed for all subjects. Then we write the joint distribution of the $r$-dimensional covariate vector $(x_{hi1}, \ldots, x_{hir})$ as

$$p(x_{hi1}, \ldots, x_{hir} | \alpha) = p(x_{hir} | x_{hi1}, \ldots, x_{hi,r-1}, x_{hi, \text{obs}}, \alpha_j) \times p(x_{hi,r-1} | x_{hi1}, \ldots, x_{hi,r-2}, x_{hi, \text{obs}}, \alpha_{r-1}) \cdots p(x_{hi1} | x_{hi, \text{obs}}, \alpha_1), \quad (3.4)$$

where $\alpha_j$ is a vector of location and scale parameters for the $j$th conditional distribution, the $\alpha_j$ are distinct, and $\alpha = (\alpha_1, \ldots, \alpha_r)$. It is important to note that (3.4) implies that a model must be specified only for those covariates that are not completely observed. When one or more covariates are completely observed for all subjects, then those covariates $x_{hi, \text{obs}}$ may be used as fixed regressor variables when the joint distribution of $(x_{hi1}, \ldots, x_{hir})$ is specified. This is helpful in limiting the number of nuisance parameters that must be estimated. Although any covariate density may be specified, certain densities, such as those in the exponential family, simplify the required Gibbs sampling procedure.

We present the following practical guidelines for modeling the covariate distribution:

1. When a covariate is completely observed, we need not specify its distribution.
2. When the missing covariates are all dichotomous, we propose modeling their joint distribution using a sequence of logistic regression models. Define $u_{j-1} = (x_{\text{mis},j-1}, \ldots, x_{\text{mis},1}, x_{\text{obs}})$. Then we specify a sequence of logistic regressions for each $p(x_{\text{mis},j} | u_{j-1}, \alpha_j)$.
3. If the missing covariates are categorical with multiple levels or are counts, models such as multinomial logistic regression models, cumulative logit models, or Poisson regression models may be used.
4. When the missing covariates are continuous and take values on the real line, we may specify a joint multivariate normal distribution. Another possibility is to use a sequence of linear regressions for each $p(x_{\text{mis},j} | u_{j-1}, \alpha_j)$. For strictly positive $x_{\text{mis},j}$, we may use a log transform and then specify a normal distribution for the transformed covariates.
5. With missing mixed covariates, we specify the distribution of the continuous covariates first, and then we specify the distribution of the categorical covariates conditional on the continuous covariates. Suppose, for example, that we have two missing covariates $x_{\text{mis}} = (x_{\text{mis},1}, x_{\text{mis},2})$, where $x_{\text{mis},1}$ is continuous and $x_{\text{mis},2}$ is categorical. Then our convention for specifying the joint distribution for the missing covariates is given by $p(x_{\text{mis}} | x_{\text{obs}}, \alpha) = p(x_{\text{mis},2} | x_{\text{mis},1}, x_{\text{obs}}, \alpha_2) \times p(x_{\text{mis},1} | x_{\text{obs}}, \alpha_1).$ So we might specify a normal distribution for $x_{\text{mis},1}$ and use a logistic regression model for $p(x_{\text{mis},2} | x_{\text{mis},1}, x_{\text{obs}}, \alpha_2)$. We note that the estimates of parameters in the cure rate model are quite robust to various orders or conventions of conditioning, so that the opposite convention (specifying the distribution of discrete covariates first, and then specifying the distribution of the continuous covariates conditional on the discrete covariates) is equally appropriate.
Ibrahim et al. (1999a) outline a general procedure for the specification of the model for the missing data mechanism. Using their proposal, we suggest modeling the missing data mechanism with a sequence of one-dimensional conditional distributions. This is an approach similar to that described for the specification of the covariate distribution. Let \( R_{hij} \) indicate whether or not the \( j \)th covariate \( x_{hij} \) is observed for subject \( i \) in cluster \( h \). In addition, we assume that the first \( r \) missing covariates are nonignorable. Then we propose the model

\[
p(R_{h11}, \ldots, R_{hir} \mid x_{hi}, \phi) = p(R_{hir} \mid R_{h11}, \ldots, R_{hi,r-1}, x_{hi}, \phi_1) \\
\times p(R_{hi,r-1} \mid R_{h11}, \ldots, R_{hi,r-2}, x_{hi}, \phi_{r-1}) \times \cdots \times p(R_{hi1} \mid x_{hi}, \phi_1),
\]

(3.5)

where \( \phi_j \) is a vector of parameters for the \( j \)th conditional distribution and \( \phi = (\phi_1, \ldots, \phi_r) \). Because each \( R_{hij} \) is dichotomous, a sequence of logistic regressions may be used for (3.5). As discussed in Ibrahim et al. (1999b), this specification greatly reduces the number of nuisance parameters that must be estimated and closely approximates a joint log-linear model for the missing data indicators.

An important consideration in the modeling of the missing data mechanism is that certain specifications may lead to inestimable parameters. The issue of estimability often arises in nonignorable response models, and care should be taken to choose a parsimonious model for the missing data mechanism. We note that the missing data mechanism need only be specified for those missing covariates that are believed to be nonignorable and that it is not necessary to specify the missing data mechanism for MAR or MCAR covariates. The one-dimensional conditional specification also facilitates efficient sampling from the conditional distribution of the missing covariates given the observed data, which is necessary in the E-step. Using logistic regressions for each one-dimensional conditional in (3.5) ensures that each \( R_{hij} \) is log-concave in \( \phi_j \) and therefore guarantees that (3.5) is log-concave in \( \phi \). This greatly eases the computations required in the E-step of the algorithm.

Along with the results of a primary analysis, a detailed sensitivity analysis of both the missing data mechanism and covariate distribution should be presented. In this manner we are able to check whether the parameter estimates of interest (i.e. the parameters of the cure rate model) depend heavily on the assumptions made about the covariate distribution and missing data mechanism. It is well known that misspecification of the missing data mechanism in particular (such as presenting a complete case analysis, which assumes MCAR missing data, when data are in fact nonignorable) can lead to misleading parameter estimates and conclusions. We suggest estimating a variety of models for the missing data mechanism that include models under MCAR, MAR, and nonignorable scenarios. We stress that the sensitivity analysis cannot be used to test a certain missing data mechanism or the adequacy of the covariate distribution. However, robustness in the parameters of the cure rate model is reassuring, and widely varying parameter estimates would suggest caution in the interpretation of results from any analysis. In any case, the investigator’s knowledge of the subject matter at hand should be the primary guide to choosing the most reasonable models for the covariate distribution and missing data mechanism.
3.3 Estimation with missing data

When some covariates are missing, the full data log-likelihood corresponding to (2.2), incorporating the
covariate distribution and (if needed) the missing data mechanism, is written

\[
l(\gamma \mid D) = \sum_{h=1}^{H} \sum_{i=1}^{n_h} v_{hi} \log(N_{hi}) + v_{hi} \log(\lambda_{khi}) + N_{hi} \log(S^*(y_{hi} \mid \lambda)) + \sum_{h=1}^{H} \sum_{i=1}^{n_h} N_{hi}(x'_{hi}\beta + w'_{hi}\beta) - \log(N_{hi}!) - \exp(x'_{hi}\beta + w'_{hi}\beta) \]
\[
+ \sum_{h=1}^{H} \frac{-q}{2} \log(2\pi) - \frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2}(b_{hi}'\Sigma^{-1}b_{hi}) + \sum_{h=1}^{H} \sum_{i=1}^{n_h} \log(p(x_{hi,\text{mis}} \mid x_{hi,\text{obs}}, \alpha)) + \log(p(R_{hi} \mid x_{hi}, \phi)),
\]

where now \( \gamma = (\beta, \lambda, \Sigma, \alpha, \phi). \) As in the complete covariate case, we implement an EM algorithm
and take the expectation of the log-likelihood with respect to the unobserved random effects, missing
covariates, and latent variables. Thus the E-step with missing covariates is given by

\[
E[I(\gamma \mid D) \mid x_{hi}, \gamma^{(t)}] = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \int_{x_{\text{mis}}, b, N} p(x_{hi,\text{mis}}, b_{hi}, N_{hi} \mid x_{hi,\text{obs}}, \gamma^{(t)}) p(x_{hi,\text{mis}}, b_{hi}, N_{hi} \mid x_{hi,\text{obs}}, \gamma^{(t)}) \]
\[
\times (v_{hi} \log(N_{hi}) + v_{hi} \log(\lambda_{khi}) + N_{hi} \log(S^*(y_{hi} \mid \lambda^{(t)})) \int x_{hi,\text{mis}} \) \(db_{hi} dN_{hi}
\]
\[
+ \sum_{h=1}^{H} \sum_{i=1}^{n_h} \int_{x_{\text{mis}}, b, N} p(x_{hi,\text{mis}}, b_{hi}, N_{hi} \mid x_{hi,\text{obs}}, \gamma^{(t)}) \]
\[
\times \{N_{hi}(x'_{hi}\beta + w'_{hi}\beta) - \log(N_{hi}!) - \exp(x'_{hi}\beta + w'_{hi}\beta)\} dx_{hi,\text{mis}} db_{hi} dN_{hi}
\]
\[
+ \sum_{h=1}^{H} \int_{x_{\text{mis}}, b, N} p(x_{hi,\text{mis}}, b_{hi}, N_{hi} \mid x_{hi,\text{obs}}, \gamma^{(t)}) \]
\[
\times \left\{-\frac{q}{2} \log(2\pi) - \frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2}(b_{hi}'\Sigma^{-1}b_{hi})\right\} dx_{hi,\text{mis}} db_{hi} dN_{hi}
\]
\[
+ \sum_{h=1}^{H} \sum_{i=1}^{n_h} \int_{x_{\text{mis}}, b, N} p(x_{hi,\text{mis}}, b_{hi}, N_{hi} \mid x_{hi,\text{obs}}, \gamma^{(t)}) \]
\[
\times \log(p(x_{hi,\text{mis}} \mid x_{hi,\text{obs}}, \alpha)) + \log(p(R_{hi} \mid x_{hi}, \phi)) dx_{hi,\text{mis}} db_{hi} dN_{hi}, \quad (3.7)
\]

where \( p(x_{hi,\text{mis}}, b_{hi}, N_{hi} \mid x_{hi}, \gamma^{(t)}) \) is the density of the random effects, missing covariates, and latent
count variables conditional on the observed data. Computational details are given in the Appendix.

3.4 Variance estimation

Variance estimation for the parameters of interest is complicated by several factors. Because the estimates
are obtained from an EM algorithm, one possibility is to use Louis’s (1982) method to calculate the
observed information matrix. However, the dimension of \( \lambda \) and the other nuisance parameters may
be large, and the variance estimates may be unstable. In addition, some error is introduced by Gibbs
maximum likelihood estimation in random effects cure rate models

A simple variance estimator with good small-sample properties was recently introduced by Goetghebeur and Ryan (2000). Following Rubin and Schenker (1991), they propose imputing data (in this case, the imputed data would be sampled random effects and sampled values of the unobserved covariates) and obtaining naive point estimates and variance estimates for the parameters of interest. Then the variance of the EM estimator is obtained as a weighted sum of the mean of the imputation variances and the empirical variance of the imputation point estimates, with weights 1 and $1 + \frac{1}{m}$, respectively, where $m$ is the number of imputations used. The variance estimation proceeds as follows.

1. Run the proposed EM algorithm until convergence, where $\hat{\gamma}$ represents the parameter estimates at convergence.
2. Fixing the parameter estimates at $\hat{\gamma}$, impute one value for each of the missing covariates. Impute one random effect for each cluster.
3. Obtain parameter estimates and their variances based on the information matrix.
4. Repeat steps 2 and 3 $m$ times.
5. Obtain the final variance estimate as the sum of the mean of the imputation variances and $(1 + \frac{1}{m})$ times the empirical variance of the imputation point estimates.

4. Analysis of Melanoma Data

We consider data from two phase III clinical trials, E1684 (Kirkwood et al., 1996) and E1690 (Kirkwood et al., 1999), conducted by ECOG. Before these trials, no adjuvant chemotherapy had been shown to have a significant impact on the survival of melanoma patients after surgery. Interferon Alpha-2b is a post-operative chemotherapy treatment used for other cancers that was believed to have the potential to make a significant impact on survival in high-risk melanoma patients as well. We consider the effect of this treatment on relapse-free survival, defined as the time from randomization until recurrence of cancer or death. Because the treatment effect seen in E1684 was larger than anticipated and was accompanied by substantial side effects due to the high dosage, ECOG began a confirmatory trial, E1690, to check the results of E1684 and to study the effect of a lower dosage of interferon. This trial was designed for the same patient population as E1684, and the high-dose interferon and observation arms in this trial were identical to those of E1684. In addition, E1690 contained a low-dose interferon arm that will not be incorporated in this analysis.

Although these data were initially analyzed using a log-rank test and the Cox (1972) regression model, these methods do not address the fact that some fraction of patients are no longer at risk of death or relapse due to melanoma after sufficient follow-up. This effect can be seen as a plateau effect in the Kaplan–Meier plots given in Figure 1, where clearly $S(\infty)$ is greater than zero. In each Kaplan–Meier plot in Figure 1, the survival curves do not approach zero but tend to stabilize in the range of 20% for observation and 40% for interferon. This effect may be further examined by looking at survival plots from a cure rate model with treatment as a covariate. These survival plots are also shown in Figure 1. The cure rate model provides a better estimate of the survival function in the tails (and hence the cure rate) than the Kaplan–Meier estimate, which may be sensitive at late event times where data are sparse.

In both studies, high-dose interferon led to greater patient survival, though statistical significance was achieved only in E1684. Investigators believed that steady improvements in the treatment and care of high-risk melanoma patients produced a cohort or time-of-entry effect in the study despite the fact that patients received exactly the same therapies over time. In order to account for this possibility, we introduce a random effect for time interval of entry onto the studies. We have 57 clusters, each containing a cohort of patients entering the study during roughly a two month time period. The cluster sizes range from four to 25 patients. To model the hazard $\lambda$, we assumed a constant hazard $\lambda_k$ in each of the five intervals.
also standardized to have mean zero and variance one. Analyses to achieve approximate normality in the distributions of the continuous covariates, which were modeled and are conditioned upon throughout the analysis as usual. We model tumor type, a dichotomous variable, using a logistic regression model of the form

\[ \phi(x_{hi}, x_{hi3}, y_{hi}, \phi_3) = \frac{\exp(\phi_{10} + \phi_{41}x_{hi4} + \phi_{42}x_{hi5} + \phi_{43}x_{hi8} + \phi_{44}y_{hi})}{1 + \exp(\phi_{10} + \phi_{41}x_{hi4} + \phi_{42}x_{hi5} + \phi_{43}x_{hi8} + \phi_{44}y_{hi})}, \]

where \( \phi_1 = (\phi_{10}, \phi_{11}, \phi_{12}, \phi_{13})' \), \( \phi_2 = (\phi_{20}, \phi_{21}, \phi_{22}, \phi_{23})' \), and \( \phi_3 = (\phi_{30}, \phi_{31}, \phi_{32}, \phi_{33})' \). This specification allows the probability of observing one covariate to depend on the possibly unobserved value of that covariate itself and the observed event time. Other missing data mechanisms are considered in the sensitivity analysis presented in Table 4.

Because these three covariates are missing, we must also specify their joint distribution. Using (3.4), we model the covariate distribution as

\[ p(x_{hi3}, x_{hi4}, x_{hi5} | x_{hi2}, x_{hi3}, x_{hi4}, \alpha) = p(x_{hi5} | x_{hi2}, x_{hi3}, x_{hi4}, \alpha_1) p(x_{hi4} | x_{hi2}, \alpha_2) p(x_{hi3} | x_{hi2}, \alpha_3), \]

where \( \alpha_1 = (\alpha_{10}, \alpha_{11}, \alpha_{12}, \alpha_{13})' \). We then model the continuous covariates Breslow thickness and size as normal random variables. We take \( x_{hi4} | x_{hi2}, x_{hi3}, \alpha_2 \sim N(\alpha_{20} + \alpha_{21}x_{hi2} + \alpha_{22}x_{hi3}, \alpha_{23}) \), where \( \alpha_2 = (\alpha_{20}, \alpha_{21}, \alpha_{22}, \alpha_{23}) \). Then we take \( x_{hi3} | x_{hi2}, \alpha_3 \sim N(\alpha_{30} + \alpha_{31}x_{hi2}, \alpha_{32}) \), where \( \alpha_3 = (\alpha_{30}, \alpha_{31}, \alpha_{32}) \).

Because of the computational intensity of the procedure, we take a burn-in of 200 Gibbs samples and use 200 samples for each missing covariate and random effect in the analysis. The convergence criterion was that the squared distance be less than 10^{-4} between the \( l \)th and \((l + 20)\)th EM iterations.

We present the results of this analysis in Table 1. We compare these not only to the results from a complete case analysis discarding all the missing subjects but also to the results from an analysis that
assumes the missing covariates are not nonignorable but MAR instead. For the MAR analysis, we use the same covariate distribution described for the nonignorable analysis and given in (4.9) but leave the missing data mechanism unspecified. We see that the complete case analysis, which is valid when the missing covariates are MCAR, finds no significant treatment benefit. Both the MAR and nonignorable EM analyses show that interferon is significantly better than observation. The MAR and nonignorable EM analyses show that advanced age and larger tumor size are detrimental towards disease-free survival and that Breslow thickness and tumor type do not have any significant effect.

It is also of interest to estimate the cure rates under the three models. Recall that the cure rate is given by exp(− exp(β x + w′ b)). If we fix the values of Breslow thickness, tumor size, and age at 0 (their average transformed values) and consider a cluster with \( h_t = 0 \), we obtain the cure rate estimates presented in Table 2. The complete case cure rates appear to be overly optimistic, estimating cure rates of roughly 40% for both treatments with not much difference between treatments. The MAR and nonignorable EM analyses yield similar estimates of the cure rates that are both somewhat lower than those in the complete case analysis. These models also show a larger difference due to treatment. According to the EM models, the cure rates for patients on interferon are roughly 10% higher than the cure rates for patients on observation.

We estimate median survival in the uncured from the relationship

\[
S^*(t) = \frac{\exp(\theta \bar{F}(t)) - \exp(-\theta)}{1 - \exp(-\theta)},
\]

where \( \bar{F}(t) = 1 - \exp(-\lambda t - s_k t) - \sum_{k=1}^{s_k} \lambda_k(s_k t - s_k t - 1), s_k t < t \leq s_k \), and \( \theta = \exp(\beta x + w' b) \). When we look at the estimated median survival times for the uncured patients in the study given in Table 3, we see again that the complete case estimates are optimistic, with estimates of median survival that are greater than the EM estimates. Using the EM analysis, we see a greater difference due to treatment than in the complete case analysis. Using the EM model that assumes the missing data are nonignorable, median survival for uncured patients on interferon is roughly one month longer than for patients on observation. In all models, there does not appear to be an effect of tumor type on median survival. Estimates of median survival from a standard Cox model would be contaminated by the inclusion of those patients who have been ‘cured’ by surgery along with the ‘uncured’ patients.

It is important to address the sensitivity of the modeling scheme to both the specification of the missing data mechanism and the specification of the covariate distribution. With this in mind, we conducted sensitivity analyses of both aspects of the model. We report the sensitivity analysis for the missing data mechanism in Table 4. For the sensitivity analysis of the missing data mechanism, we used the same covariate distribution, \( p(x_{hi5} | x_{hi2}, x_{hi3}, x_{hi4}, \alpha_1) p(x_{hi4} | x_{hi2}, x_{hi3}, \alpha_2) p(x_{hi3} | x_{hi2}, \alpha_3) \), in (4.9) and varied the missing data mechanism. We consider several different parametrizations for the missing data mechanism.

- **MD1**: mechanism left unspecified (MAR);
- **MD2**: \( p(R_4 | x_4, y, \phi_2) \) (Breslow thickness and type MAR but size may be nonignorable);
- **MD3**: \( p(R_4 | x_4, y, \phi_2) p(R_4 | x_3, y, \phi_3) \) (type MAR but Breslow thickness and size may be nonignorable);
- **MD4**: \( p(R_5 | x_5, y, \phi_1) p(R_4 | x_4, y, \phi_2) p(R_3 | x_3, y, \phi_3) \) (Breslow thickness, size, and type may be nonignorable; same as mechanism given in (4.8));
- **MD5**: \( p(R_5 | x_5, y, R_3, R_4, \phi_1) p(R_4 | x_4, y, R_3, \phi_2) p(R_3 | x_3, y, \phi_3) \) (Breslow thickness, size, and type may be nonignorable, and missing data indicators may be correlated).

In Table 4 we see that the model is generally robust to variations in the specification of the missing data mechanism. The estimates of the treatment effect are virtually identical in all five models considered, and the models all agree that interferon has a significant benefit. In addition, all models agree that advanced age and larger tumor size are related to poor outcomes and that tumor type has no significant effect. The
### Table 1. Results for melanoma data

<table>
<thead>
<tr>
<th></th>
<th>CC</th>
<th>MAR</th>
<th>NI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cure rate model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.14 (0.07)</td>
<td>0.21 (0.06)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.25 (0.06)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Trt</td>
<td>−0.12 (0.09)</td>
<td>−0.28 (0.07)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.28 (0.07)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Age</td>
<td>0.05 (0.04)</td>
<td>0.09 (0.04)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.08 (0.04)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Breslow</td>
<td>0.00 (0.05)</td>
<td>0.02 (0.04)</td>
<td>0.06 (0.04)</td>
</tr>
<tr>
<td>Size</td>
<td>0.06 (0.04)</td>
<td>0.08 (0.04)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.08 (0.03)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Type</td>
<td>0.08 (0.10)</td>
<td>0.06 (0.09)</td>
<td>−0.02 (0.08)</td>
</tr>
<tr>
<td><strong>Random effects sigma</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.18 (0.02)</td>
<td>0.11 (0.01)</td>
<td>0.10 (0.01)</td>
</tr>
<tr>
<td><strong>Covariate distribution</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breslow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–</td>
<td>0.00 (0.04)</td>
<td>0.08 (0.04)</td>
</tr>
<tr>
<td>Age</td>
<td>–</td>
<td>0.08 (0.04)</td>
<td>0.07 (0.04)</td>
</tr>
<tr>
<td>σ&lt;sub&gt;bres&lt;/sub&gt;</td>
<td>–</td>
<td>0.99 (0.05)</td>
<td>1.05 (0.06)</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–</td>
<td>−0.01 (0.04)</td>
<td>−0.25 (0.04)</td>
</tr>
<tr>
<td>Age</td>
<td>–</td>
<td>0.12 (0.04)</td>
<td>0.15 (0.04)</td>
</tr>
<tr>
<td>Breslow</td>
<td>–</td>
<td>0.19 (0.04)</td>
<td>0.18 (0.04)</td>
</tr>
<tr>
<td>σ&lt;sub&gt;size&lt;/sub&gt;</td>
<td>–</td>
<td>0.96 (0.05)</td>
<td>1.07 (0.06)</td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–</td>
<td>−0.73 (0.10)</td>
<td>−0.58 (0.10)</td>
</tr>
<tr>
<td>Age</td>
<td>–</td>
<td>−0.11 (0.10)</td>
<td>−0.12 (0.09)</td>
</tr>
<tr>
<td>Breslow</td>
<td>–</td>
<td>−1.05 (0.11)</td>
<td>−0.96 (0.10)</td>
</tr>
<tr>
<td>Size</td>
<td>–</td>
<td>0.28 (0.10)</td>
<td>0.16 (0.09)</td>
</tr>
<tr>
<td><strong>Hazard</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00–0.50</td>
<td>0.55 (0.05)</td>
<td>0.61 (0.05)</td>
<td>0.61 (0.05)</td>
</tr>
<tr>
<td>0.50–1.00</td>
<td>0.59 (0.07)</td>
<td>0.63 (0.06)</td>
<td>0.63 (0.06)</td>
</tr>
<tr>
<td>1.00–2.50</td>
<td>0.53 (0.06)</td>
<td>0.50 (0.05)</td>
<td>0.50 (0.05)</td>
</tr>
<tr>
<td>2.50–5.00</td>
<td>0.44 (0.08)</td>
<td>0.49 (0.08)</td>
<td>0.49 (0.08)</td>
</tr>
<tr>
<td>5.00–10.00</td>
<td>0.43 (0.22)</td>
<td>0.54 (0.24)</td>
<td>0.54 (0.24)</td>
</tr>
<tr>
<td><strong>Missing data mechanism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breslow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–</td>
<td>–</td>
<td>6.21 (1.17)</td>
</tr>
<tr>
<td>Breslow</td>
<td>–</td>
<td>–</td>
<td>−2.90 (0.67)</td>
</tr>
<tr>
<td>Time</td>
<td>–</td>
<td>–</td>
<td>−1.73 (0.65)</td>
</tr>
<tr>
<td>Bres*time</td>
<td>–</td>
<td>–</td>
<td>1.22 (0.35)</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–</td>
<td>–</td>
<td>2.70 (0.28)</td>
</tr>
<tr>
<td>Size</td>
<td>–</td>
<td>–</td>
<td>1.63 (0.23)</td>
</tr>
<tr>
<td>Time</td>
<td>–</td>
<td>–</td>
<td>0.04 (0.19)</td>
</tr>
<tr>
<td>Size*time</td>
<td>–</td>
<td>–</td>
<td>0.04 (0.16)</td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–</td>
<td>–</td>
<td>2.30 (0.20)</td>
</tr>
<tr>
<td>Type</td>
<td>–</td>
<td>–</td>
<td>−0.46 (0.37)</td>
</tr>
<tr>
<td>Time</td>
<td>–</td>
<td>–</td>
<td>0.30 (0.13)</td>
</tr>
<tr>
<td>Type*time</td>
<td>–</td>
<td>–</td>
<td>−0.62 (0.24)</td>
</tr>
</tbody>
</table>

<sup>a</sup> p < 0.05 in cure rate model.
Table 2. Cure rates, calculated for subjects with average values of Breslow thickness, size, and age, and random effect estimate $\hat{b} = 0$

<table>
<thead>
<tr>
<th>Model</th>
<th>Superficial spreading tumors</th>
<th>Other tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interferon (%)</td>
<td>Observation (%)</td>
</tr>
<tr>
<td>Complete case</td>
<td>40</td>
<td>36</td>
</tr>
<tr>
<td>MAR</td>
<td>37</td>
<td>27</td>
</tr>
<tr>
<td>Nonignorable</td>
<td>39</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 3. Median survival (in years) for the uncured, calculated for subjects with average values of Breslow thickness, size, and age, and random effect estimate $\hat{b} = 0$

<table>
<thead>
<tr>
<th>Model</th>
<th>Superficial spreading tumors</th>
<th>Other tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interferon</td>
<td>Observation</td>
</tr>
<tr>
<td>Complete case</td>
<td>0.82</td>
<td>0.78</td>
</tr>
<tr>
<td>MAR</td>
<td>0.78</td>
<td>0.68</td>
</tr>
<tr>
<td>Nonignorable</td>
<td>0.79</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table 4. Sensitivity analysis for melanoma missing data mechanism. The model used in the primary analysis is denoted CD3-MD4

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
</tr>
<tr>
<td>CD3-MD1</td>
<td>0.21 (0.06)</td>
</tr>
<tr>
<td>CD3-MD2</td>
<td>0.23 (0.06)</td>
</tr>
<tr>
<td>CD3-MD3</td>
<td>0.20 (0.06)</td>
</tr>
<tr>
<td>CD3-MD4</td>
<td>0.25 (0.06)</td>
</tr>
<tr>
<td>CD3-MD5</td>
<td>0.25 (0.06)</td>
</tr>
</tbody>
</table>

$p < 0.05$.

The estimate of the coefficient of Breslow thickness does vary depending on the missing data mechanism, so we judge the effect of Breslow thickness with caution. Due to the similarity of the MAR and nonignorable models, we feel there is evidence that the covariates are MAR and that an analysis using MD1, which allows us to leave the missing data mechanism unspecified, is appropriate. The results of our sensitivity analysis for the missing data mechanism highlight the importance of trying a variety of models rather than choosing one model for the missing data mechanism and ignoring the need for a sensitivity analysis.

In addition, we conducted a sensitivity analysis of the covariate distributions. In this case, we used the missing data mechanism $P(R_5 \mid x_3, y, \phi_1) P(R_4 \mid x_4, y, \phi_2) P(R_3 \mid x_5, y, \phi_3)$, given in (4.8), and we varied the covariate parametrizations while still using logistic regression for tumor type and linear regression for Breslow thickness and tumor size. The results of this sensitivity analysis are presented in Table 5. The following covariate parametrizations were considered in this sensitivity analysis:

- CD1: $p(x_5 \mid \alpha_1) p(x_4 \mid \alpha_2) p(x_3 \mid \alpha_1)$ (smallest possible model for Breslow thickness, size, and type);
- CD2: $p(x_5 \mid x_3, x_4, \alpha_1) p(x_4 \mid x_3, \alpha_2) p(x_3 \mid \alpha_1)$;
- CD3: $p(x_5 \mid x_2, x_3, x_4, \alpha_1) p(x_4 \mid x_2, x_3, \alpha_2) p(x_3 \mid x_2, \alpha_1)$ (described in (4.9) and used in primary analysis as well as sensitivity analysis of missing data mechanism);
Table 5. Sensitivity analysis for melanoma covariate distribution. The model used in the primary analysis is denoted CD3-MD4

<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept (SE)</th>
<th>Treatment (SE)</th>
<th>Age (SE)</th>
<th>Breslow (SE)</th>
<th>Size (SE)</th>
<th>Type (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD1-MD4</td>
<td>0.30 (0.06)a</td>
<td>-0.29 (0.07)a</td>
<td>0.08 (0.04)a</td>
<td>0.01 (0.04)</td>
<td>0.07 (0.04)b</td>
<td>-0.15 (0.08)b</td>
</tr>
<tr>
<td>CD2-MD4</td>
<td>0.24 (0.06)a</td>
<td>-0.28 (0.07)a</td>
<td>0.08 (0.04)a</td>
<td>0.06 (0.04)</td>
<td>0.10 (0.04)a</td>
<td>-0.01 (0.09)</td>
</tr>
<tr>
<td>CD3-MD4</td>
<td>0.25 (0.06)a</td>
<td>-0.28 (0.07)a</td>
<td>0.08 (0.04)a</td>
<td>0.06 (0.04)</td>
<td>0.08 (0.03)a</td>
<td>-0.02 (0.08)</td>
</tr>
<tr>
<td>CD4-MD4</td>
<td>0.23 (0.06)a</td>
<td>-0.27 (0.07)a</td>
<td>0.08 (0.04)a</td>
<td>0.07 (0.04)b</td>
<td>0.08 (0.03)a</td>
<td>0.03 (0.09)</td>
</tr>
</tbody>
</table>

a \( p < 0.05 \), b \( p < 0.10 \).

- CD4: \( p(x_5 \mid x_1, x_2, x_3, x_4, \alpha_1) \ p(x_4 \mid x_1, x_2, x_3, \alpha_2) \ p(x_3 \mid x_1, x_2, \alpha_1) \).

With respect to inference about the treatment effect, the models all agree that treatment is beneficial and have treatment effect estimates that are almost exactly the same. However, it appears that CD1-MD4, the smallest possible covariate model, is not sufficient to model the covariate distribution. In fact, with this model we see a significant improvement in survival for superficial spreading tumors. This effect is not seen in any other models fit to the data. So one should exercise caution in the specification of the covariate distribution and try several models in order to ensure that the distribution is appropriately modeled.

5. DISCUSSION

We propose a method of estimation in cure rate models with random effects. In addition, we give an extension of the model that allows us to obtain unbiased parameter estimates in the presence of possibly nonignorable (or MAR) missing covariates. The methodology introduced is computationally feasible and is used in the analysis of two ECOG melanoma clinical trials. The proposed analysis (1) attempts to characterize the type and impact of missing tumor characteristics, providing some evidence that the missing data may not be MCAR, (2) shows evidence of a cohort or time of entry effect, and (3) allows estimation of the probability of cure for these high-risk patients. A detailed sensitivity analysis was conducted to assess the dependence of the results obtained on the modeling assumptions. We evaluated models that assumed the missing data were MCAR, MAR, or nonignorable. As a whole, this particular analysis was not very sensitive to parametric assumptions about the missing data mechanism or distributions of the missing covariates when we assumed the missing data were MAR or nonignorable. However, we stress the importance of conducting a sensitivity analysis whenever one accounts for missing data. The results of our analysis lead us to conclude that the missing data are most likely not nonignorable but in fact are most likely MAR and therefore that a complete case analysis is most likely not appropriate for this particular dataset and would, in this case, lead the investigators to conclusions different from those obtained in the MAR and nonignorable models.

One issue that warrants further consideration is accurate estimation of the cure rate. The estimates can be unstable when follow-up is insufficient or when too few events occur, and in such cases the model itself may not be identifiable. Chen et al. (1999) develop a Bayesian model with informative prior distributions that allows better estimation of cure rates in such settings. This model could be adapted for the random effects cure rate structure proposed here. Although the cure rate estimates were relatively robust in the various sensitivity analyses for the melanoma data, these estimates did vary slightly along with the parameter estimates in these analyses.

In addition, one might argue against the incorporation of the random effect in the analysis. Although incorporating a random effect makes the analysis more difficult and time-consuming, the missing data and
Suppose \( \lambda_{k_i} \) is the estimate of \( \lambda \) in the interval \( (s_{k_i-1}, s_{k_i}] \) containing the event time \( y_{hi} \). (We provide a formula for estimating \( \lambda \) after describing the Monte Carlo EM estimation scheme.) Then from (2.1), we have

\[
f^*(y_{hi} | \lambda) = \lambda_{k_i} \exp \left\{ - \left[ \lambda_{k_i} (y_{hi} - s_{k_i-1}) + \sum_{g=1}^{k_i-1} \lambda_g (s_g - s_{g-1}) \right] \right\}
\]

and

\[
S^*(y_{hi} | \lambda) = \exp \left\{ - \left[ \lambda_{k_i} (y_{hi} - s_{k_i-1}) + \sum_{g=1}^{k_i-1} \lambda_g (s_g - s_{g-1}) \right] \right\}.
\]

**APPENDIX: TECHNICAL DETAILS**

**Formulae for \( f^* \) and \( S^* \)**

Suppose \( \lambda_{k_i} \) is the estimate of \( \lambda \) in the interval \( (s_{k_i-1}, s_{k_i}] \) containing the event time \( y_{hi} \). (We provide a formula for estimating \( \lambda \) after describing the Monte Carlo EM estimation scheme.) Then from (2.1), we have

\[
f^*(y_{hi} | \lambda) = \lambda_{k_i} \exp \left\{ - \left[ \lambda_{k_i} (y_{hi} - s_{k_i-1}) + \sum_{g=1}^{k_i-1} \lambda_g (s_g - s_{g-1}) \right] \right\}
\]

and

\[
S^*(y_{hi} | \lambda) = \exp \left\{ - \left[ \lambda_{k_i} (y_{hi} - s_{k_i-1}) + \sum_{g=1}^{k_i-1} \lambda_g (s_g - s_{g-1}) \right] \right\}.
\]

To evaluate the E-step with no missing covariates given in (2.3), we will use the Monte Carlo EM algorithm of Wei and Tanner (1990). We write the joint density of the random effects and latent counts conditional on the observed data as

\[
p(b_h | x_{hi}, y^{(l)}) = p(N_{hi} | x_{hi}, b_h, y^{(l)}) p(b_h | x_{hi}, y^{(l)}).
\]

It can be shown that \( [N_{hi} | x_{hi}, b_h, y^{(l)}] \sim V_{hi} + y_{hi}, \) where \( V_{hi} \sim \text{Poisson}(S^*(y_{hi} | \lambda) \exp(x_{hi}^T y^{(l)} + w_{hi}^T b_h)) \), and

\[
p(b_h | x_{hi}, y^{(l)}) \propto p(b_h | \Sigma^{(l)}) \prod_{j=1}^{m_h} \left\{ \exp\left(x_{hi}^T \beta^{(l)} + w_{hi}^T b_h \right) f^*(y_{hi} | \lambda^{(l)}) \right\}^{\nu_{hi}}
\]

\[
\times \exp(- \exp(x_{hi}^T \beta^{(l)} + w_{hi}^T b_h) [1 - S^*(y_{hi} | \lambda^{(l)})]).
\]

We note that the density in (A.1) has an attractive form for Gibbs sampling. As long as \( p(b_h | \Sigma) \) is log-concave, then \( p(b_h | x_{hi}, y^{(l)}) \) is also log-concave. This enables us to use the adaptive rejection algorithm of Gilks and Wild (1992) in our sampling scheme when we sample the unobserved random effects.

Suppose that for the \( h \)th cluster, we take a sample of size \( m_h \), \( (b_h^{(1)}, b_h^{(2)}, \ldots, b_h^{(m_h)}) \) from \( p(b_h | x_{hi}, y^{(l)}) \) using the Gibbs sampler. Noting that the terms \( \log(N_{hi}) \) and \( \log(N_{hi}!\) are constants with respect to the estimation of \( \gamma \), we see that the log-likelihood is essentially linear in \( N_{hi} \), and thus the E-step may be simplified by replacing \( N_{hi} \) with its conditional expectation. This expectation, which depends on \( b_h^{(j)} \), the \( j \)th sampled value of \( b_h \), is given by

\[
N_{hi}^{(l+1)} = E(N_{hi} | b_h^{(j)} , y^{(l)}) = S^*(y_{hi} | \lambda^{(l)}) \exp(x_{hi}^T \beta^{(l)} + w_{hi}^T b_h)
\]
Note that we can break down the E-step as
\[ Q(\gamma | \gamma^{(l)}) = \sum_{h=1}^H \frac{1}{m_h} \sum_{j=1}^{m_h} \sum_{i=1}^{n_h} N_{hi}^{(l+1)} (x_{hi} \beta + w_{hi} b_{h}^{(j)}) - \exp(x_{hi} \beta + w_{hi} b_{h}^{(j)}) \]
\[ + \sum_{h=1}^H \frac{1}{m_h} \sum_{j=1}^{m_h} \sum_{i=1}^{n_h} v_{hi} \log(\lambda_{ki}) + N_{hi}^{(l+1)} \log(S^*(y_{hi} | \lambda^{(l)})) \]
\[ + \sum_{h=1}^H \frac{1}{m_h} \sum_{j=1}^{m_h} -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2} (b_{h}^{(j)} - \Sigma^{-1} b_{h}^{(j)})^2, \]
where the terms \( \log(N_{hi}) \) and \( \log(N_{hi}^{(l)}) \), constants with respect to the estimation of \( \gamma \), have been dropped. Note that we can break down the E-step as \( Q(\gamma | \gamma^{(l)}) = Q^{(1)}(\beta | \gamma^{(l)}) + Q^{(2)}(\lambda | \gamma^{(l)}) + Q^{(3)}(\Sigma | \gamma^{(l)}) \), where we may use Newton–Raphson to obtain \( \beta^{(l+1)} \), the maximizer of \( Q^{(1)}(\beta | \gamma^{(l)}) \). The estimate of \( \Sigma^{(l+1)} \) is straightforward to calculate, and \( \lambda_k^{(l+1)} \) is given by
\[ \lambda_k^{(l+1)} = \left\{ \sum_{h_{k-1} < h < h_k} v_{hi} \right\} \times \left\{ \sum_{h_{k-1} < h < h_k} \left[ \frac{1}{m_h} \sum_{j=1}^{m_h} N_{hi}^{(l+1)} \right] (y_{hi} - s_{h1}) + \sum_{h_{k-1} < h < h_k} \left[ \frac{1}{m_h} \sum_{j=1}^{m_h} N_{hi}^{(l+1)} \right] (s_k - s_{h1}) \right\}^{-1}, \]
for \( k = 1, \ldots, K \). We see that the estimate of \( \lambda \) exists as long as at least one event occurs in each interval \( (s_{h1}, s_k) \).

**Monte Carlo EM when some covariates are missing**

To evaluate the E-step in (3.7) with missing covariates, consider \( p(x_{hi, mis}, b_{h} | N_{hi} | x_{hi}, \gamma^{(l)}) = p(N_{hi} | x_{hi}, b_{h}, \gamma^{(l)}) p(x_{hi, mis} | x_{hi, obs}, \gamma^{(l)}) \), where \( [N_{hi} | x_{hi}, b_{h}, \gamma^{(l)}] \sim \mathcal{V}_{hi} + v_{hi} \) and \( v_{hi} \sim \text{Poisson}(S^*(y_{hi} | \lambda^{(l)}) \exp(x_{hi} \beta^{(l)} + w_{hi} b_{h})) \). In addition,
\[ p(x_{hi, mis}, b_{h} | x_{hi, obs}, \gamma^{(l)}) \propto p(b_{h} | \Sigma) \times \prod_{i=1}^{n_h} p(R_{hi} | x_{hi}, \phi^{(l)}) p(x_{hi, mis} | x_{hi, obs}, \alpha^{(l)}) \]
\[ (\exp(x_{hi} \beta^{(l)} + w_{hi} b_{h}) f^*(y_{hi} | \lambda^{(l)}) \right)^{v_{hi}} \exp(- \exp(x_{hi} \beta^{(l)} + w_{hi} b_{h})[1 - S^*(y_{hi} | \lambda^{(l)})]) \} \]. \hspace{1cm} (A.2)

To evaluate (3.7), we use the Monte Carlo EM algorithm of Wei and Tanner (1990) in much the same manner as with complete covariate data. Samples will be obtained from \( [x_{hi, mis}, b_{h} | x_{hi, obs}, \gamma^{(l)}] \) using the Gibbs sampler (Gelfand and Smith, 1990) along with the adaptive rejection algorithm of Gilks and Wild (1992). The quantity in (A.2) has an attractive form for sampling because it is log-concave provided that we choose log-concave densities for \( p(b_{h} | \Sigma) \), \( p(x_{hi, mis} | x_{hi, obs}, \alpha) \), and \( p(R_{hi} | x_{hi}, \phi) \).

For each subject, let \( (x_{hi}^{(j)}, b_{h}^{(j)}), j = 1, \ldots, m_h \) contain the \( j \)th sampled values of random effects and the missing covariates, with \( x_{hi}^{(j)} = (x_{hi, obs}, x_{hi, mis}) \). Then we have \( N_{hi}^{(l+1)} = E(N_{hi} | x_{hi}^{(j)}, b_{h}^{(j)}, \gamma^{(l)}) = \).
Maximum likelihood estimation in random effects cure rate models

\[ S^*(y_{hi} \mid \lambda^{(l)}) \exp(x_{hi}^{(l)} \beta^{(l)}) + w_{hi}^{(l)} b_h^{(l)} + v_{hi}, \]

which leads to the E-step

\[ Q(y \mid y^{(l)}) = \sum_{h=1}^{H} \frac{1}{m_h} \sum_{j=1}^{m_h} \sum_{i=1}^{n_h} N_{hij}^{(l+1)} (x_{hi}^{(l)} \beta + w_{hi}^{(l)} b_h^{(l)}) - \exp(x_{hi}^{(l)} \beta + w_{hi}^{(l)} b_h^{(l)}) \]

\[ + \sum_{h=1}^{H} \frac{1}{m_h} \sum_{j=1}^{m_h} \sum_{i=1}^{n_h} v_{hi} \log(\lambda_{ki}) + \lambda_{hi}^{(l+1)} \log(S^*(y_{hi} \mid \lambda^{(l)})) \]

\[ + \sum_{h=1}^{H} \frac{1}{m_h} \sum_{j=1}^{m_h} \sum_{i=1}^{n_h} \log(p(x_{hi,\text{mis}}^{(j)} \mid \alpha)) + \log(p(R_{hi} \mid x_{hi}^{(j)} \mid \phi)) \]

\[ + \sum_{h=1}^{H} \frac{1}{m_h} \sum_{j=1}^{m_h} \sum_{i=1}^{n_h} \frac{q}{2} \log(2\pi) - \frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2} (b_h^{(j)} \Sigma^{-1} b_h^{(j)}). \]

We break down the E-step as \( Q(y \mid y^{(l)}) = Q^{(1)}(\beta \mid y^{(l)}) + Q^{(2)}(\lambda \mid y^{(l)}) + Q^{(3)}(\Sigma \mid y^{(l)}) + Q^{(4)}(\alpha \mid y^{(l)}) + Q^{(5)}(\phi \mid y^{(l)}), \)

and we may use Newton–Raphson to obtain \( \lambda^{(l+1)}, \alpha^{(l+1)}, \) and \( \phi^{(l+1)}. \)

We estimate \( \lambda^{(l+1)} \) and \( \Sigma^{(l+1)} \) as before.

To summarize, the proposed EM algorithm proceeds as follows:

- obtain an initial estimate of \( y \), say \( y^{(0)} \) = \( (\beta^{(0)}, \lambda^{(0)}, \Sigma^{(0)}, \alpha^{(0)}, \phi^{(0)}) \). In the absence of any prior information regarding the structure of \( \Sigma \), one may take \( \Sigma^{(0)} = I_{q \times q} \) as a starting value;
- sample the random effects, \( b_a \), and the missing covariates, \( x_{hi,\text{mis}} \);
- compute the conditional expectations \( N_{hij}^{(l)} \);
- at the \( (l + 1)\)th EM iteration, update the parameter estimates to obtain \( y^{(l+1)} \);
- iterate until convergence.

REFERENCES


Maximum likelihood estimation in random effects cure rate models


[Received May 24, 2001; revised August 27, 2001; accepted for publication August 28, 2001]