The Canadian Journal of Statistics
Vol. 29, No. 3, 2001, Pages 421–436
La revue canadienne de statistique

Survival analysis with long-term survivors and partially observed covariates

Meehyung CHO, Nathaniel SCHENKER, Jeremy M. G. TAYLOR & Dongliang ZHUANG

Key words and phrases: Cure model; general location model; Gibbs sampling; logistic regression; missing data; mixture model; radiation therapy; Weibull model.

MSC 2000: Primary 62P10; secondary 62F15.

Abstract: The authors describe a method for fitting failure time mixture models that postulate the existence of both susceptibles and long-term survivors when covariate data are only partially observed. Their method is based on a joint model that combines a Weibull regression model for the susceptibles, a logistic regression model for the probability of being a susceptible, and a general location model for the distribution of the covariates. A Bayesian approach is taken, and Gibbs sampling is used to fit the model to the incomplete data. An application to clinical data on tonsil cancer and a small Monte Carlo study indicate potential large gains in efficiency over standard complete-case analysis as well as reasonable performance in a variety of situations.

Analyse des survivants à long terme dans les situations où les covariables ne sont pas toutes observées

Résumé : Les auteurs décrivent une méthode d'ajustement de modèles de mélanges pour des temps de défaillance dans lesquels on postule l'existence de cas à risque et de survivants à long terme dans les situations où les covariables ne sont pas toutes observées. Leur méthode exploite à la fois un modèle de régression de Weibull pour les cas à risque, un modèle de régression logistique pour la probabilité d'être à risque, et un modèle de localisation général pour la loi des covariables. Une approche bayésienne est adoptée et l'échantillonneur de Gibbs est employé pour ajuster le modèle à des données incomplètes. Une application à des données cliniques sur le cancer des amygdales et une petite étude de Monte-Carlo illustrent les gains d'efficacité substantiels qu'il est possible de réaliser en utilisant cette méthode plutôt qu'en restreignant l'analyse aux cas pour lesquels toute l'information concernant les covariables est disponible.

1. INTRODUCTION

This paper discusses mixture models for failure time data that postulate a population with two types of individuals: susceptibles and long-term survivors. The susceptibles are at risk of developing the event under consideration, and the event would be observed with certainty if complete follow-up were possible. The long-term survivors will never experience the event, however. Failure time mixture models have been applied in a wide range of disciplines, such as oncology (Taylor 1995; Peng, Dear & Denham 1998; Sy & Taylor 2000; Zhuang *et al.* 2000), sociology (Yamaguchi 1992), criminology (Maller & Zhao 1996) and engineering reliability (Meeker 1987).

The assumption that some individuals will never experience the event is sensible in certain situations. For example, Taylor (1995) observed that only between 5% and 50% of patients with tumours of the head and neck will experience local recurrences after radiation therapy. The remaining patients will not have recurrences because all of the tumour cells have been killed by the radiation. The existence of long-term survivors suggests modeling both the probability of being a susceptible and the failure time distribution for susceptibles. The probability of being a susceptible is typically modeled by a logistic regression. Examples of regression models that have been used for the failure times of susceptibles include a Weibull model (Farewell 1982, 1986), a proportional hazards model (Kuk & Chen 1992; Sy & Taylor 2000; Zhuang *et al.* 2000), an accelerated failure time model (Yamaguchi 1992), a Kaplan–Meier type of nonparametric model (Taylor 1995), and a generalized F model (Peng, Dear & Denham 1998).

One complication in fitting failure time mixture models is that for each censored failure time, it is not known whether the individual is a long-term survivor or a susceptible. An additional complication occurs when covariate data are incomplete. While efficient approaches such as maximum likelihood estimation have been used to handle the censoring problem (e.g., Larson & Dinse 1985), "complete-case analysis" has typically been used to handle missing covariates. Complete-case analysis uses only those subjects with complete covariate information, discarding any individual with a missing covariate value. Depending on the missing-data mechanism and the extent of missing data, this approach can lead to both biased estimates and loss of efficiency.

In this paper, we describe a method for fitting failure time mixture models to data with incomplete covariate information that is based on jointly modeling the outcome and covariate data. Such joint modeling is generally useful in problems with missing covariates and incomplete response information (e.g., Little & Schluchter 1985; Faucett, Schenker & Elashoff 1998). We combine a standard failure time mixture model with a general location model (Olkin & Tate 1961) for the covariates. For computational ease, we take a Bayesian approach and use Gibbs sampling (Geman & Geman 1984; Gelfand & Smith 1990) to fit the model. In a study described in Zhuang *et al.* (2000), we applied our approach to reanalyze data on head and neck cancer from Dubray *et al.* (1996). Here, we provide details of the method, apply it to data from a study of tonsil cancer (Withers *et al.* 1995), and evaluate it in a Monte Carlo study.

Our models and analytical method are described in Section 2. In Section 3, we present our analysis of tonsil cancer data, and we compare our method to complete-case analysis. A Monte Carlo study of our method under a variety of conditions is summarized in Section 4. Possible extensions of our method as well as alternative methods are discussed in Section 5.

2. MODELS AND ANALYTICAL METHOD

Our joint model includes three submodels: a model for the covariates (continuous and/or categorical), a model for the probability of being a susceptible, and a model for the distribution of failure times among susceptibles. Note that a model for the distribution of failure times among long-term survivors is not needed, since the long-term survivors are assumed not to fail. Alternatively, the distribution of failure times among long-term survivors may be thought of as degenerate at infinity.

2.1. Notation.

Let (T, D, X, Z) denote the random variables of interest for each subject, defined as follows: T is the failure time (which is finite only if the subject is susceptible); D is a binary indicator of susceptibility, where $D = 1$ indicates that the subject is a susceptible and $D = 0$ indicates that the subject is a long-term survivor; X is a p-dimensional vector of continuous covariates; and Z is a q-dimensional vector of categorical covariates.

Typically, X and Z are subject to missingness; moreover, T and D are observed only when the subject experiences the event during follow-up. When a subject is censored, D is unobserved, and it is only known that T is greater than the censoring time.

2.2. Models.

The joint density function of (T, D, X, Z) can be expressed as

$$\begin{aligned} f(t, d, x, z|\theta, \beta, \eta, \pi) &= \{f(t|D = 1, z, x; \theta)P(D = 1|z, x; \beta)\}^d \\ &\quad \times P(D = 0|z, x; \beta)^{1-d} f(x|z; \eta) f(z|\pi), \end{aligned}$$

where $f(t|D = 1, z, x; \theta)$ is the density function of the failure time of a susceptible given the covariates; $P(D = 0|z, x; \beta)$ and $P(D = 1|z, x; \beta)$ are the probabilities of being a long-term survivor and a susceptible, respectively, given the covariates; $f(x|z; \eta)$ is the density function of the continuous covariates given the categorical covariates; and $f(z|\pi)$ is the density function of

the categorical covariates. The main interest is in the parameters β and θ, with η and π regarded as nuisance parameters.

We use the general location model for the distribution of covariates, corresponding to $f(x|z;\eta)\ f(z|\pi)$. The general location model was proposed for mixtures of categorical and continuous variables by Olkin & Tate (1961) and extended by Krzanowski (1980, 1982) to include constraints among the parameters. Maximum likelihood methods for fitting the general location model to incomplete data were developed by Little & Schluchter (1985). Jointly modeling the failure time and covariates, Cho & Schenker (1999) and Meng & Schenker (1999) proposed methods to handle missing covariates in survival analysis with the accelerated failure time model. For a recent discussion of the general location model, see Schafer (1997, Chapter 9).

Suppose that the jth element of Z, $j = 1, \ldots, q$, has k_j levels. Then the possible values of Z define a contingency table with $K = k_1 \times \cdots \times k_q$ cells. Assume that a subject falls into cell ℓ with probability π_ℓ, where $\pi_1 + \cdots + \pi_K = 1$. Thus, if we define an indicator for the cell of the contingency table in which a subject belongs, the indicator will have a multinomial distribution with probabilities $\pi_1, \ldots, \pi_K$. Conditional on the subject belonging to cell ℓ, X is assumed to have a p-variate normal distribution with mean vector μ_ℓ and variance-covariance matrix Σ_ℓ, $\ell = 1, \ldots, K$. The general location model traditionally specifies a separate mean vector μ_ℓ for each cell but a common variance-covariance matrix Σ across the cells, and places no restrictions on the cell probabilities except for the obvious restriction that $\pi_1 + \cdots + \pi_K = 1$. We generalize the traditional model by assuming a separate variance-covariance matrix Σ_ℓ for each cell.

As is traditional in failure time mixture models, we assume that the distribution of D follows a logistic regression model with

$$P(D = 1|z, x; \beta) = \frac{\exp(\beta_0 + z'\beta_1 + x'\beta_2)}{1 + \exp(\beta_0 + z'\beta_1 + x'\beta_2)}, \tag{1}$$

where $\beta = (\beta_0, \beta_1', \beta_2')'$ is a parameter vector with β_1 and β_2 being vectors.

We consider the Weibull regression model for the failure time T conditional on $D = 1$ [corresponding to $f(t|D = 1, z, x; \theta)$], with hazard function of the form

$$h(t|D = 1; \lambda, \tau) = \lambda\tau(\lambda t)^{\tau - 1},$$

where $\lambda > 0$ is the scale parameter and $\tau > 0$ is the shape parameter. To allow the dependence of the failure time on the covariates, we model λ as a function of the covariates,

$$\lambda(z, x; \alpha) = \exp(\alpha_0 + z'\alpha_1 + x'\alpha_2),$$

where $\alpha = (\alpha_0, \alpha_1', \alpha_2')'$ is a parameter vector with α_1 and α_2 being vectors. Thus, given $D = 1$, T has hazard function

$$h(t|D = 1, z, x, ; \theta) = \exp(\alpha_0 + z'\alpha_1 + x'\alpha_2)\tau\{\exp(\alpha_0 + z'\alpha_1 + x'\alpha_2)t\}^{\tau - 1}, \tag{2}$$

survival function

$$S(t|D = 1, z, x; \theta) = \exp[\{-\exp(\alpha_0 + z'\alpha_1 + x'\alpha_2)t\}^{\tau}], \tag{3}$$

and density function

$$f(t|D = 1, z, x; \theta) = h(t|D = 1, z, x, ; \theta)S(t|D = 1, z, x, ; \theta), \tag{4}$$

where $\theta = \{\tau, \alpha\}$.

With regard to the missing-data and censoring mechanisms, we make some standard assumptions, as described for example in Schluchter & Jackson (1989). Specifically, we assume that the mechanism resulting in missing predictors is ignorable, which implies that the missing predictors are either missing completely at random or missing at random (Rubin 1976); that censoring is random and noninformative (Kalbfleisch & Prentice 1980, p. 40); and that the censoring distribution does not depend on any predictors that are missing.

2.3. Prior distributions.

As mentioned in Section 1, we use Bayesian methods to fit our joint model to data. The parameters of the joint model are $\pi = (\pi_1, \ldots, \pi_K)$ and $\eta = (\mu_1, \ldots, \mu_K, \Sigma_1, \ldots, \Sigma_K)$ for the general location model for Z and X; β for the logistic regression for D; and $\theta = \{\tau, \alpha\}$ for the Weibull regression model for T.

For the situation in which no covariate values are missing so that it is not necessary to model Z and X, Chen, Ibrahim & Sinha (1999) showed that if an improper uniform prior distribution is specified for β, then the posterior distribution of (β, θ) is improper regardless of the prior distribution specified for θ. To avoid such problems, we specify diffuse but proper prior distributions for most of the parameters in our model, including β and θ. Another approach to avoiding the problem pointed out by Chen, Ibrahim & Sinha (1999) is to assume that all subjects that are censored after the last observed failure time are long-term survivors, as suggested by Taylor (1995).

All of the parameters in our model, except for the components of π, are assumed to be a priori independent. For each scalar component of α, we specify a $N(0, v_\alpha)$ prior distribution with a large value of v_α. Similarly, the prior distribution for each scalar component of β is $N(0, v_\beta)$ with a large value of v_β. For τ, we specify a gamma prior distribution with shape parameter a and scale parameter b (i.e., mean a/b and variance a/b^2), such that the distribution has a large variance. For (μ_ℓ, Σ_ℓ), $\ell = 1, \ldots, K$, we specify independent Jeffreys prior distributions with densities $f(\mu_\ell, \Sigma_\ell) \propto |\Sigma_\ell|^{-(p+1)/2}$. Finally, for $\pi_1, \ldots, \pi_K$, the prior distribution is Dirichlet with each hyperparameter equal to u, where u is small. Specific values for the hyperparameters of our prior distributions are given in Sections 3 and 4.

2.4. Method of analysis.

2.4.1. Overview.

We approximate the posterior distribution of the joint model parameters by using Gibbs sampling (Geman & Geman 1984; Gelfand & Smith 1990) to obtain approximate draws from the posterior distribution. Gibbs sampling is an iterative procedure in which each unobserved quantity is drawn in turn from its full conditional distribution given the observed data and the most recent draws of the other unobserved quantities. When a full conditional distribution is only known up to a normalizing constant, the Metropolis-Hastings algorithm (Hastings 1970; Smith & Roberts 1993) within the Gibbs sampler is used in obtaining draws. After convergence, the procedure results in draws from the joint posterior distribution of the unobserved quantities. In the context of our problem, the unobserved quantities are the model parameters, the missing covariates, and the values of D for the censored cases. The full conditional distributions used in the Gibbs sampling procedure are described in Section 2.4.2, and modifications for when cell counts associated with the general location model are small are discussed in Section 2.4.3. To monitor convergence of the Gibbs sampling procedure, we examine time series plots of the parameter draws.

2.4.2. Full conditional distributions for Gibbs sampling.

If the values of Z, X, and D were completely observed, the likelihood for the parameters of our joint model could be expressed as

$$\prod_{i=1}^{n} [\{f(t_i^*|D_i = 1, z_i, x_i; \theta)\}^{c_i} \{S(t_i^*|D_i = 1, z_i, x_i; \theta)\}^{1-c_i} P(D_i = 1|z_i, x_i; \beta)]^{d_i} \times \{P(D_i = 0|z_i, x_i; \beta)\}^{1-d_i} f(x_i|z_i; \eta) f(z_i|\pi), \tag{5}$$

where for subject i ($1 \le i \le n$), the z_i, x_i, and d_i are the observed values of Z, X, and D; t_i^* is the minimum of the failure time and the censoring time; and $c_i = 1$ if t_i^* is the failure time and $c_i = 0$ if t_i^* is the censoring time. Note that the factor

$$[\{f(t_i^*|D_i = 1, z_i, x_i; \theta)\}^{c_i} \{S(t_i^*|D_i = 1, z_i, x_i; \theta)\}^{1-c_i} P(D_i = 1|z_i, x_i; \beta)]^{d_i}$$

in expression (5) is equal to 1 when $d_i = 0$. Thus, only subjects with $d_i = 1$ (i.e., susceptibles) contribute information about θ.

The full conditional density of each quantity drawn in the Gibbs sampling procedure (i.e., a parameter, a missing covariate value, or a value of D for a censored subject) is proportional to the product of the prior distributions and the likelihood (5). All factors not involving the quantity being drawn are treated as constants.

In particular, under the prior distributions specified in Section 2.3, the full conditional distributions are as follows.

(a) The distribution of the cell probabilities $\pi_1, \ldots, \pi_K$ for the general location model is Dirichlet with density proportional to ${\pi_1}^{n_1+u-1} \times \cdots \times {\pi_K}^{n_K+u-1}$, where n_ℓ, $\ell = 1, \ldots, K$, is the number of subjects belonging to the ℓth cell of the contingency table.

(b) The distribution of (μ_ℓ, Σ_ℓ), $\ell = 1, \ldots, K$, can be factorized into the distribution of Σ_ℓ, which is inverse-Wishart with degrees of freedom $n_\ell - 1$ and scale S_ℓ, and the distribution of μ_ℓ given Σ_ℓ, which is $N(\overline{x}_\ell, \Sigma_\ell/n_\ell)$, where $\overline{x}_\ell$ is the sample mean of X within cell ℓ, $S_\ell = \sum_{i \in B_\ell} (x_i - \overline{x}_\ell)(x_i - \overline{x}_\ell)'$ is the matrix of sums of squares and cross products within cell ℓ, and $B_\ell = \{i \mid \text{subject } i \text{ belongs in cell } \ell\}$.

In the following expressions, we express the remaining full conditional distributions in terms of $f(t|D = 1, z, x; \theta)$ and $S(t|D = 1, z, x; \theta)$, which are defined in equations (2)–(4), $P(D = 1|z, x; \beta)$, which is defined in equation (1), $P(D = 0|z, x; \beta) = 1 - P(D = 1|z, x; \beta)$, and $f(x|z; \eta)$, which is a p-variate normal density with mean vector μ_ℓ and covariance matrix Σ_ℓ, where ℓ indicates the cell in the general location model corresponding to the categorical covariate values z.

(i) The density of α is proportional to

$$\exp\{-\alpha'\alpha/(2v_\alpha)\} \prod_{i=1}^{n} \{f(t_i^*|D_i = 1, z_i, x_i; \theta)\}^{c_i d_i} \{S(t_i^*|D_i = 1, z_i, x_i; \theta)\}^{(1-c_i)d_i}.$$

(ii) The density of τ is proportional to

$$\tau^{a-1} e^{-\tau/b} \prod_{i=1}^{n} \{f(t_i^*|D_i = 1, z_i, x_i; \theta)\}^{c_i d_i} \{S(t_i^*|D_i = 1, z_i, x_i; \theta)\}^{(1-c_i)d_i}.$$

(iii) The density of β is proportional to

$$\exp\{-\beta'\beta/(2v_\beta)\} \prod_{i=1}^{n} \{P(D_i = 1|z_i, x_i; \beta)\}^{d_i} \{P(D_i = 0|z_i, x_i; \beta\}^{1-d_i}.$$

In the remaining expressions, we omit the subscript i for simplicity.

(iv) For each subject with missing components of Z, the probability that $Z = z(\ell)$ is proportional to

$$\pi_\ell [\{f(t^*|D = 1, z(\ell), x; \theta)\}^c \{S(t^*|D = 1, z(\ell), x; \theta)\}^{1-c} P(D = 1|z(\ell), x; \beta)]^d \times \{P(D = 0|z(\ell), x; \beta)\}^{1-d} f(x|z(\ell); \eta),$$

for $\ell \in R$, where $z(\ell)$ represents the value of Z corresponding to cell ℓ of the contingency table, and R is the set of possible cells for the subject based on the observed components of Z.

(v) For each subject with missing components of X, the density of X is proportional to

$$[\{f(t^*|D=1,z,x;\theta)\}^c\{S(t^*|D=1,z,x;\theta)\}^{1-c}P(D=1|z,x;\beta)]^d$$
$$\times\{P(D=0|z,x;\beta)\}^{1-d}f(x|z;\eta),$$

for all values of x such that the components corresponding to non-missing components of X are fixed at the observed values.

(vi) For each censored subject, the probability that $D=1$ is given by

$$\frac{P(D=1|z,x;\beta)S(t^*|D=1,z,x;\theta)}{P(D=1|z,x;\beta)S(t^*|D=1,z,x;\theta)+P(D=0|z,x;\beta)}.$$

The full conditional densities for α, β, τ, and the missing components of X are not densities of standard distributions, and they are only known up to normalizing constants. Hence, for α, β, τ, and the missing components of X, we use the Metropolis-Hastings algorithm (Hastings 1970; Smith & Roberts 1993) within the Gibbs sampler in obtaining draws.

2.4.3. Handling small cell counts.

At each iteration of Gibbs sampling, if a cell, say cell ℓ, has fewer than $p+1$ observations in it after the missing values in Z are drawn, then the full conditional distribution of (μ_ℓ,Σ_ℓ) is improper. Therefore, when the contingency table based on the completely observed values of Z in a data set has any cells with fewer than $p+1$ observations in them, the model and Gibbs sampling procedure described above may need modification. This situation did not occur in our application and Monte Carlo study (Sections 3 and 4), but we now briefly discuss possible modifications.

Potential problems due to small cell counts can be avoided either by using proper prior distributions for (μ_ℓ,Σ_ℓ), $\ell=1,\ldots,C$, or by imposing restrictions on the parameters in the general location model as discussed in Krzanowski (1980, 1982), Little & Schluchter (1985), and Schafer (1997, Section 9.3). One type of restriction would be to revert to the traditional assumption (see Section 2.2) that the variance-covariance matrix of X has a constant value, say Σ, across the cells. Under this restriction and the Jeffreys prior, the full conditional distribution of Σ is still inverse-Wishart, but the degrees of freedom become $n-C$ and the scale becomes the pooled matrix of sums of squares and cross products (Schafer 1997, Section 9.2.4). Thus, the requirement of at least $p+1$ observations per cell at each iteration of Gibbs sampling is replaced by the less stringent requirements of at least $p+C$ observations in the entire sample and at least one observation in each cell.

Potential problems with estimating the cell means due to empty cells can be alleviated by placing restrictions on the means. Specifically, instead of specifying a saturated multivariate regression with X as the outcome variable and all of the main effects and interactions of Z as predictors, as is implicitly done when the cell means are allowed to vary freely, we may specify an unsaturated model with only certain main effects and interactions included. See Schafer (1997, Section 9.3.3) for details of the full conditional distributions under such a model when a common variance-covariance matrix across the cells is also assumed.

In the situation in which it is known that none of the subjects in the study belongs to cell ℓ, an alternative type of prior restriction would be to treat the contingency table as having a "structural zero" in cell ℓ, that is, to assume that $\pi_\ell=0$. This would result in leaving cell ℓ out of the Gibbs sampling procedure, and thus in not making inferences about (μ_ℓ,Σ_ℓ).

3. APPLICATION TO TONSIL CANCER DATA

Data were obtained from a collaborative retrospective study of external beam radiation therapy for treating squamous cell carcinomas of the tonsil (Withers *et al.* 1995). The nine institutions

in the study used a variety of dose-fractionation schemes, and the purpose of the study was to investigate the importance of treatment defined by total dose and overall treatment duration. The outcome of interest, local recurrence of cancer, was experienced during follow-up by 208 of the 676 patients for whom data were collected.

Among the covariate data collected, we consider tumour stage, node stage, age, sex, a measure of hemoglobin, total dose, and overall treatment duration. Hemoglobin was missing for 235 subjects, node stage was missing for 3 subjects, and age was missing for 1 subject. (Seven of the missing values of hemoglobin resulted from deletion of recorded values that were far outside the plausible range presumably due to recording error.) Tumour stage, sex, total dose, and overall treatment duration were observed completely.

The data were analyzed by Sy & Taylor (2000) using a mixture model and maximum likelihood estimation. The failure times for susceptibles were modelled using a Weibull model and a Cox proportional hazards model. Sy & Taylor (2000) included the covariates tumour stage, node stage, age, sex, total dose, and overall treatment duration in the model, but individuals with missing data were excluded from the analysis. Thus, 672 patients were included in the analysis.

We reanalyze the data using our joint modeling approach with the logistic/Weibull mixture model. The covariate hemoglobin was excluded from the analyses of Sy & Taylor (2000) because over one third of the subjects had no hemoglobin measurements. Our approach allows us to include hemoglobin while retaining all of the information on the other covariates. We are interested in hemoglobin because it has been found to be a prognostic factor for local recurrence of head and neck cancer following radiation therapy (Dubray *et al.* 1996; Zhuang *et al.* 2000), with higher levels of hemoglobin associated with lower risk of local recurrence.

Our analysis includes three categorical covariates (node stage, tumour stage, and sex) that define a contingency table with sixteen cells. There are also four continuous covariates (treatment duration, total dose, hemoglobin, and age), which are assumed in the general location model to have a multivariate normal distribution within each cell of the contingency table.

The values of the hyperparameters used in our prior distributions (see Section 2.3) were $v_\alpha = 2500$, $v_\beta = 2500$, $a = 1$, $b = 0.1$, and $u = 1.5$. In the Gibbs sampling procedure, we discarded the draws in the first 600,000 iterations for complete-case analysis and the first 800,000 iterations for joint modeling. We then saved every fifteenth sample (to reduce autocorrelation) until we had 25,000 draws with which to approximate each posterior distribution.

Figure 1 displays time series plots of the draws from the joint modeling analysis, for the regression coefficients emphasized in the remainder of this section.

Table 1 shows the results obtained from the complete-case and joint modeling analyses. The point estimates (posterior medians) from the two methods are comparable, especially relative to the posterior standard deviations. Thus the pattern of missing data does not result in a major change in point estimates when the incomplete cases are included in the analysis via joint modeling.

Zhuang (1999) compared the distributions of variables among the complete cases with those among the incomplete cases and found little difference. This is consistent with missingness completely at random, a mechanism under which both the complete-case and joint modeling analyses are valid (Cho & Schenker 1999).

The joint modeling analysis shows that the strongest evidence of effects (posterior 95% intervals that do not include 0) is for predictors in the logistic regression part of the model. In particular, in the logistic regression, there is strong evidence that the coefficients for tumour stage and treatment duration are positive and that the coefficient of total dose is negative. There is also a consistent ordering with respect to tumour stage. In the Weibull regression, there is only strong evidence that the coefficient of age is negative. These findings are consistent with those described by Sy & Taylor (2000). Thus, including hemoglobin in the model has little influence on the estimated relationships between the outcomes and the other predictors.

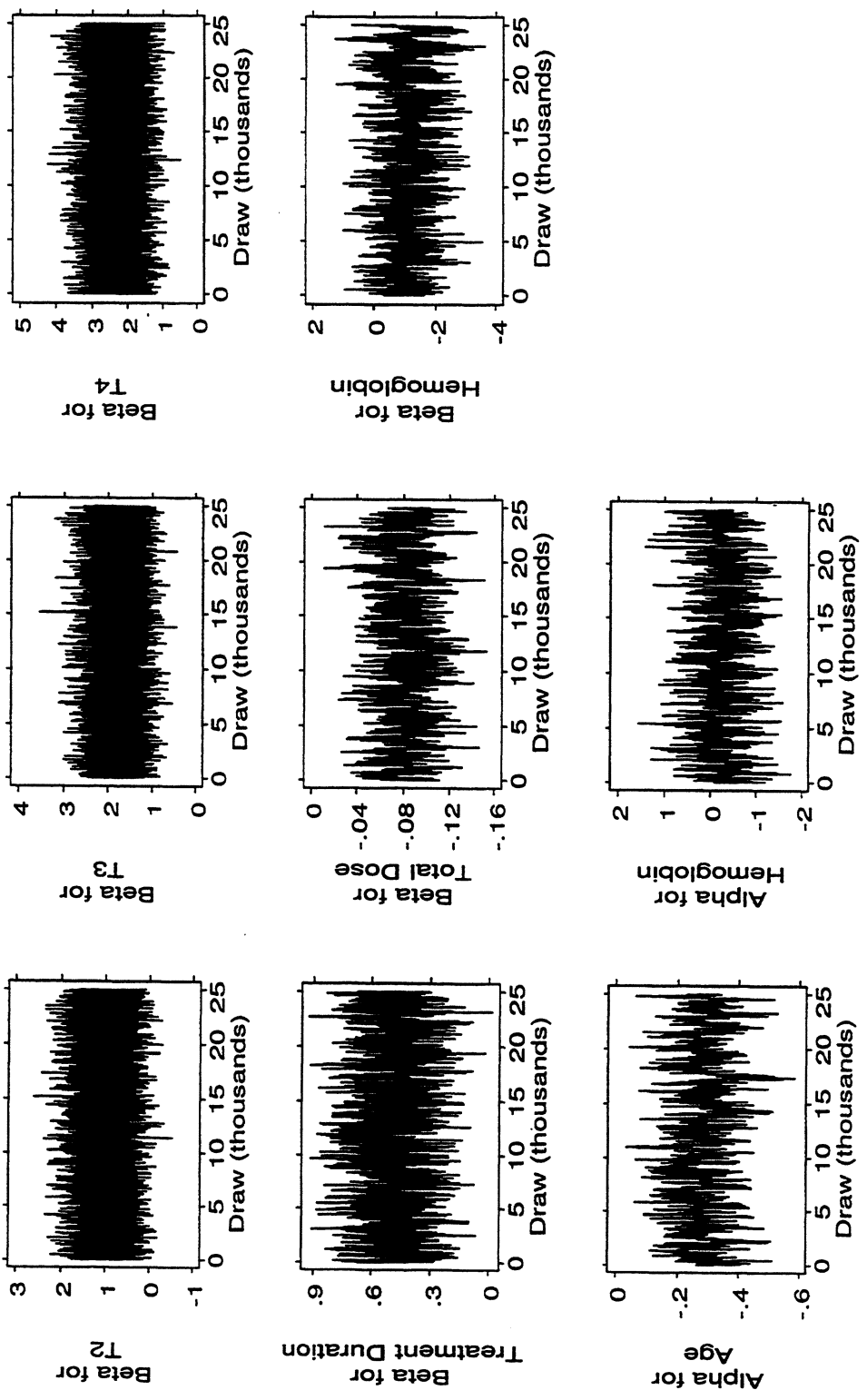

FIGURE 1: Selected time series plots from the joint modeling analysis. ("Beta" denotes a logistic regression coefficient and "Alpha" denotes a Weibull regression coefficient.)

TABLE 1: Results from fitting the logistic/Weibull mixture model to the tonsil cancer data: complete-case analysis versus joint modeling.

	Complete-case analysis ($n = 439$)			Joint modeling ($n = 676$)		
	Est.	SD	95% Interval	Est.	SD	95% Interval
Logistic regression						
Intercept	2.065	1.928	(-1.482, 5.983)	1.438	1.397	(-1.268, 4.044)
Node stage[a]	0.383	0.251	(-0.099, 0.885)	0.428	0.203	(0.034, 0.832)
Tumour stage:						
T2[b]	0.847	0.417	(0.076, 1.701)	0.987	0.371	(0.299, 1.756)
T3[c]	1.673	0.403	(0.923, 2.507)	1.782	0.358	(1.117, 2.524)
T4[d]	1.864	0.519	(0.874, 2.916)	2.327	0.448	(1.493, 3.232)
Treatment duration[e]	0.562	0.180	(0.240, 0.947)	0.481	0.127	(0.235, 0.735)
Total dose[f]	-0.092	0.026	(-0.148, -0.047)	-0.082	0.019	(-0.120, -0.047)
Hemoglobin[g]	-1.047	0.842	(-2.709, 0.556)	-1.113	0.668	(-2.385, 0.243)
Age[h]	0.056	0.126	(-0.184, 0.313)	0.108	0.096	(-0.075, 0.298)
Sex[i]	0.294	0.287	(-0.263, 0.867)	0.299	0.228	(-0.140, 0.754)
Weibull regression						
Intercept	3.056	1.355	(0.302, 5.637)	2.466	1.012	(0.514, 4.439)
Node stage	0.275	0.199	(-0.109, 0.679)	0.263	0.150	(-0.024, 0.562)
Tumour stage:						
T2	-0.565	0.361	(-1.236, 0.188)	-0.576	0.300	(-1.125, 0.049)
T3	-0.268	0.351	(-0.905, 0.470)	-0.249	0.292	(-0.770, 0.373)
T4	0.465	0.397	(-0.302, 1.269)	0.290	0.330	(-0.331, 0.971)
Treatment duration	-0.075	0.102	(-0.283, 0.118)	-0.020	0.068	(-0.158, 0.110)
Total dose	-0.005	0.016	(-0.039, 0.028)	-0.004	0.012	(-0.028, 0.019)
Hemoglobin	-0.329	0.549	(-1.318, 0.834)	-0.236	0.465	(-1.089, 0.746)
Age	-0.326	0.102	(-0.528, -0.127)	-0.279	0.074	(-0.430, -0.138)
Sex	-0.015	0.213	(-0.415, 0.434)	-0.059	0.166	(-0.371, 0.277)

[a]1 if positive; 0 if negative, [b]1 if stage 2; 0 if stage 1, [c]1 if stage 3; 0 if stage 1, [d]1 if stage 4; 0 if stage 1, [e]unit = 10 days, [f]unit = 1 Gray, [g]unit = 10 g/dL, [h]unit = 10 years, [i]1 if male; 0 if female.

The estimated coefficients of hemoglobin are negative in both the logistic regression and Weibull regression models, which is consistent with Dubray *et al.* (1996) and Zhuang *et al.* (2000). Moreover, as in Zhuang *et al.* (2000), the evidence that hemoglobin is associated with susceptibility is stronger than the evidence that hemoglobin is associated with failure time for susceptibles. For example, for joint modeling, the approximate posterior probability (not shown in Table 1) that the coefficient of hemoglobin is negative is 0.935 for the logistic regression, but only 0.693 for the Weibull regression. Although the evidence regarding hemoglobin in the logistic regression is inconclusive, as demonstrated by the posterior intervals covering both positive and negative values, the approximate posterior probability that the coefficient is negative is

higher for joint modeling (0.935) than for complete-case analysis (0.895). This is consistent with increased precision due to incorporating all of the data.

The posterior standard deviations in Table 1 also indicate that the joint modeling method is more efficient than the complete-case analysis. Meng & Schenker (1999) suggested that the relative efficiency of two methods could be approximated roughly by the ratio of the percentages of covariate values used by the two methods. For the seven covariates considered in our analysis, the percentage of values used by the joint modeling analysis was $\{(7)(676) - 235 - 3 - 1\}/\{(7)(676)\} = 95\%$, whereas the percentage of values used by complete-case analysis was $439/676 = 65\%$. Thus, the approximation of Meng & Schenker (1999) suggests that the relative efficiency in our example is roughly $95\%/65\% = 1.46$. It follows that the posterior standard deviations from complete-case analysis are expected to be roughly $\sqrt{1.46} = 1.21$ times as large as those from joint modeling. The ratios of posterior standard deviations in Table 1 range from 1.16 to 1.50, in rough accord with the approximation.

4. MONTE CARLO STUDY

To investigate the properties of our method further, we performed a small Monte Carlo study under the logistic/Weibull mixture model.

4.1. Comparison of joint modeling with complete-case analysis.

4.1.1. Design.

One hundred data sets of size $n = 400$ were simulated under the modeling assumptions of Section 2.2, with a single Bernoulli (0 or 1) covariate Z and a single continuous covariate X, and with the failure time regression model being the Weibull model. The joint modeling and complete-case analyses were applied to each data set. The values of the hyperparameters used in our prior distributions (see Section 2.3) were $v_\alpha = 2500$, $v_\beta = 225$, $a = 1$, $b = 0.1$, and $u = 1.5$. In the Gibbs sampling procedure, we discarded the first 30,000 to 80,000 iterations, depending on the appearance of the time series plots. We then saved every fifteenth sample until we had 5000 draws with which to approximate each posterior distribution. From each analysis, posterior medians and nominal 95% posterior intervals were obtained for the parameters of interest. The performance of the joint modeling and complete-case analyses across the 100 data sets was summarized.

Data for each subject were generated as follows. A value of Z was generated from a Bernoulli distribution with probability of success ($Z = 1$) equal to $\pi_1 = 0.4$. A value of X was then sampled from a normal distribution, either $N(\mu_0, \sigma_0^2)$ or $N(\mu_1, \sigma_1^2)$, depending on the value of Z, where $(\mu_0, \sigma_0^2) = (10, 1)$ and $(\mu_1, \sigma_1^2) = (5, 1)$. An indicator D was then generated from a Bernoulli distribution with probability of success specified by the logistic regression model (1), with $\beta = (2, 0.75, -0.2)'$. If the subject was susceptible ($D = 1$), a value of the failure time T was then generated from the Weibull regression model described by expressions (2)–(4), with $\tau = 2$ and $\alpha = (2, -1, -0.5)'$. If the subject was a long-term survivor ($D = 0$), no failure time was generated. Censoring of the outcomes was generated as follows. For each subject, a random censoring variable Y was generated from an exponential distribution with a mean of 400 days. If the subject was a susceptible, then the observed failure or censoring time was $T^* = \min(T, Y, \Delta)$, where Δ was the length of the follow-up (200 days in this simulation); if $T^* < T$, D was treated as missing. If the subject was a long-term survivor, then the censoring time was equal to $\min(Y, \Delta)$, and D was treated as missing.

With the distributions for data and censoring that were used in this simulation, 38% of the subjects were expected to be censored due to a loss to the follow-up or termination of the study. Among the censored subjects, 95% were expected to be long-term survivors. Of these long-term survivors, 61% were expected to be censored at the end of follow-up with the remaining 39% censored due to loss to follow-up. Of the expected 5% of censored subjects who were

susceptibles, virtually all were expected to be censored due to loss to follow-up. Missing data in the covariates was generated as completely at random, with 20% missingness for each covariate. Thus, 64% of subjects were expected to have complete covariate data.

4.1.2. Results.

Table 2 displays the mean estimate, the standard deviation of the estimate, and the coverage rate of the 95% posterior interval for each parameter in the logistic and Weibull regression models, for complete-case analysis and for joint modeling. Also shown is the ratio of the mean squared error for complete-case analysis to the mean squared error for joint modeling.

Both methods of analysis yield nearly unbiased estimates as well as coverage rates that are close to the nominal level. The standard deviations of the estimates from joint modeling are all lower than those from complete-case analysis, however. This is also reflected in the ratios of mean squared errors. The approximation of Meng & Schenker (1999) (see Section 3) suggests a relative efficiency of roughly $80\%/64\% = 1.25$. All of the ratios of mean squared errors are close to this value.

TABLE 2: Monte Carlo results (based on 100 trials) for the logistic/Weibull mixture model: complete-case analysis versus joint modeling.

		Complete-case analysis			Joint modeling			
Parameter	True value	Mean	SD	Coverage rate (%)	Mean	SD	Coverage rate (%)	$\frac{MSE_C}{MSE_J}$
Logistic regression								
β_0	2	2.038	1.314	96	2.062	1.154	95	1.293
β_1	0.75	0.738	0.704	95	0.788	0.621	93	1.282
β_2	-0.2	-0.199	0.131	96	-0.208	0.117	94	1.248
Weibull regression								
α_0	2	2.019	0.410	94	2.008	0.371	94	1.223
α_1	-1	-1.014	0.216	97	-1.007	0.192	94	1.273
α_2	-0.5	-0.500	0.041	96	-0.500	0.037	94	1.240
τ	2	2.000	0.115	95	1.993	0.102	95	1.278

NOTES: $n = 400$; 20% missingness for each covariate; random component of censoring is exponentially distributed with mean of 400 days and follow-up of 200 days.

4.2. Performance of joint modeling under other conditions.

To explore the performance of our methods under a variety of situations, we carried out four simulations, each with one condition changed from the set of conditions used in Section 4.1. Specifically, we considered more missing data (40% missingness of each covariate instead of 20%), smaller sample size ($n = 100$ instead of $n = 400$), heavier censoring (exponential distribution for Y with mean of 100 days instead of 400 days), and non-normal X ($11 - X$ distributed as exponential with mean 1 if $Z = 0$, $X - 4$ distributed as exponential with mean 1 if $Z = 1$, so that the means and variances of X are the same as in the normal case). The first three changes

still satisfy our modeling assumptions, whereas the fourth (non-normality) represents a violation of our assumptions. Table 3 gives the results for 100 Monte Carlo trials.

TABLE 3: Monte Carlo results (based on 100 trials) for the logistic/Weibull mixture model: joint modeling with various changes in the conditions.

		More missing data (40% for each covariate)			Smaller sample ($n = 100$)		
Parameter	True value	Mean	SD	Coverage rate (%)	Mean	SD	Coverage rate (%)
Logistic regression							
β_0	2	2.022	1.361	95	2.512	2.599	93
β_1	0.75	0.801	0.782	94	0.635	1.355	92
β_2	-0.2	-0.197	0.147	94	-0.249	0.258	94
Weibull regression							
α_0	2	2.014	0.441	94	2.089	0.812	94
α_1	-1	-1.019	0.225	93	-1.054	0.461	94
α_2	-0.5	-0.504	0.038	94	-0.509	0.081	94
τ	2	1.995	0.115	95	2.038	0.199	95

		Heavier censoring*			Non-normal X (exponential)		
Parameter	True value	Mean	SD	Coverage rate (%)	Mean	SD	Coverage rate
Logistic regression							
β_0	2	2.065	1.160	93	2.205	1.389	92
β_1	0.75	0.792	0.627	92	0.691	0.729	91
β_2	-0.2	-0.207	0.120	94	-0.217	0.139	93
Weibull regression							
α_0	2	2.015	0.392	94	1.927	0.491	83
α_1	-1	-1.011	0.195	93	-0.952	0.231	87
α_2	-0.5	-0.502	0.038	94	-0.491	0.051	89
τ	2	1.991	0.110	95	1.989	0.106	93

* Random component exponentially distributed with mean 100 days.

When our modeling assumptions are satisfied, the performance of our method is reasonable in the variety of situations that we consider. The coverage rates of the 95% posterior intervals are at least 92% in all cases, suggesting approximate validity of inferences in all of the situations. Comparison with Table 2 shows that the posterior standard deviations in Table 3 are larger to varying degrees. This is to be expected under the changes that we consider (more missing data, smaller sample size, heavier censoring), since they imply losses of information compared to the situation considered in Section 4.1. The estimates for the Weibull regression model appear nearly unbiased under the conditions that we consider. The estimates for the logistic regression model have simulated biases of varying sizes. All of the simulated biases are small relative to the simulated standard deviations, however. Moreover, the simulated biases are all within two Monte Carlo standard errors of zero, suggesting that they could be explained by Monte Carlo error.

In the case of non-normal X, all of the simulated biases are small, but the coverage rates for the Weibull regression coefficients are well below the nominal level of 95%. Even with our small number (100) of Monte Carlo trials, these low coverage rates cannot be explained by Monte Carlo error alone. Such results are expected, since the assumptions of our joint model no longer obtain. In contrast, the complete-case analysis yielded coverage rates (not shown in the table) of 91%, 93%, and 94% for α_0, α_1, and α_2, respectively, suggesting less sensitivity to non-normality. Interestingly, however, the ratios of mean squared errors (comparing complete-case analysis with the joint modeling analysis) range from 1.04 to 1.79, all favouring the joint modeling approach. Moreover, the largest ratios are for the Weibull regression coefficients. For further discussion of the assumptions required for validity of the joint modeling methods and complete-case analysis, see Cho & Schenker (1999).

The results for non-normal X suggest that, with our joint modeling approach, the use of transformations based on examining empirical covariate distributions could improve performance. Alternatively, if a covariate appears highly non-normal, then its values could be grouped into several levels to construct a closely related categorical covariate. Note, however, that if missingness in the covariate values is not completely at random, then it may be difficult to assess whether the assumption of normality is reasonable.

5. DISCUSSION

5.1. Adaptations and generalizations of our joint model.

It should be possible to adapt or generalize our joint model in many ways, of which we mention a few. If strong prior beliefs exist, they could be incorporated into the model; this could include the use of more informative prior distributions for model parameters.

Although we use a Weibull regression model for the failure times of susceptibles in this paper, the use of other failure time regression models is possible. For example, Zhuang *et al.* (2000) used a proportional hazards model with a piecewise constant baseline hazard. To modify the Gibbs sampling algorithm for use with an alternative failure time regression model, it is necessary to replace expressions (2)–(4) with expressions corresponding to the alternative failure time regression model and to incorporate the replacements in the full conditional distributions of Section 2.4.2.

One of the main features of the mixture model approach is that it allows separate consideration of the effects of a covariate on the probability of susceptibility (via the logistic regression component) and on the distribution of failure times for susceptibles (via the failure time regression component), giving two coefficients for the covariate. This was illustrated by our analysis of tonsil cancer data in Section 3. However, the use of the same covariates in the two components of the mixture model has the potential to lead to high correlations between the respective coefficient estimates, with associated problems in interpretation and convergence. This did not appear to be a problem in the joint modeling analysis of tonsil cancer data, for which the estimated correlations (from Gibbs sampling) between the Weibull regression coefficients and the corresponding logistic regression coefficients ranged from -0.26 to -0.08. Nevertheless, it is straightforward

to use different subsets of covariates in the two components. This can be desirable because it reduces the number of parameters in the model, and because it minimizes difficulties due to correlations between estimated coefficients. Development of techniques for model selection in the context of failure time mixture models is an important topic for future research.

Finally, the failure time mixture model that we use can be viewed as a special case of a frailty model (see, e.g., Keiding, Andersen & Klein 1997). In contrast to typical frailty models, our model has a "frailty variable" D that is binary and partially observed, and we model the relationship of D to selected covariates. Nevertheless, the relationship of our model to frailty models suggests that it might be possible to extend our joint modeling approach to fit various frailty models when covariate values are missing.

5.2. Some alternative models and methods of estimation.

We have chosen to formulate our model using a mixture approach to mimic the known curative potential of radiation therapy for head and neck cancer. Such mixture models are known to be potentially problematic, however, due to near non-identifiability conditions (Taylor 1995). An alternative approach would be not to use a mixture formulation but rather to use a survival model in which long-term survivors are accounted for via a hazard function that decreases appropriately to a value near zero at some time point. Chen, Ibrahim & Sinha (1999) proposed a related approach in which the population follows a model with a proportional hazards structure and a survival function that approaches a nonzero limit as time approaches infinity. Such approaches can have computational and modeling advantages over the mixture approach. Unlike mixture modeling, however, they do not allow separate interpretation of the effects of the covariates on the probability of susceptibility and on the failure time distribution for susceptibles via two sets of regression coefficients (see Section 5.1).

With regard to estimation for our joint model, there are alternatives to our approach to handling missing data in the covariates. One such alternative is maximum likelihood estimation via the EM algorithm (Dempster, Laird & Rubin 1977). For our problem, we found the E step of the EM algorithm to be analytically intractable. Rather than resorting to numerical integration, a Monte Carlo E step, or some other approximation, we chose to use a Bayesian/Gibbs sampling approach with diffuse prior distributions for ease of computation. An added benefit of Gibbs sampling is that assessments of variability (e.g., posterior standard deviations) can be calculated directly from the Gibbs samples. Given large-sample results on the similarity of maximum likelihood and Bayesian procedures (e.g., Cox & Hinkley 1974, Chapter 10), we would expect the maximum likelihood approach and our approach to yield similar inferences in many situations. In fact, for the complete cases in the tonsil cancer data, with hemoglobin omitted from the covariates, Zhuang (1999) compared our approach with a Weibull model to the results of Sy & Taylor (2000) obtained using maximum likelihood estimation. The two approaches yielded results that were quite similar.

Another alternative approach to handling missing data in the covariates is to use multiple imputation (Rubin 1987). Since multiple imputations of the missing data are a by-product of Gibbs sampling (Barnard, Rubin & Schenker 1998), it would be straightforward to multiply impute the missing covariate values using our Gibbs sampling procedure. If the completed data sets were then analyzed using the same mixture model for the outcomes that is used in the Gibbs sampling procedure, we would expect the resultant multiple-imputation inferences to be similar to inferences from our Gibbs sampling approach. This is due to the fact that multiple-imputation methods approximate the posterior distribution of the parameters of interest. Unlike our approach, multiple imputation also allows other models for the outcomes to be fitted once the missing covariate values have been imputed. Caution needs to be exercised when this is done, however, because the use of inconsistent models at the imputation and analysis stages may lead to invalid inferences (Meng 1994).

ACKNOWLEDGEMENTS

Support for this research was provided in part by grants CA64235 and CA72495 from the National Cancer Institute while Schenker, Taylor, and Zhuang were with the Dept. of Biostatistics at the University of California, Los Angeles.

REFERENCES

J. Barnard, D. B. Rubin & N. Schenker (1998). Multiple imputation methods. In *Encyclopedia of Biostatistics*, Volume 4 (P. Armitage and T. Colton, eds). Wiley, Chichester, pp. 2772–2780.

M.-H. Chen, J. G. Ibrahim & D. Sinha (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94, 909–919.

M. Cho & N. Schenker (1999). Fitting the log-F accelerated failure time model with incomplete covariate data. *Biometrics*, 55, 380–387.

D. R. Cox & D. V. Hinkley (1974). *Theoretical Statistics*. Chapman & Hall, London.

A. P. Dempster, N. M. Laird & D. B. Rubin (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39, 1–22.

B. Dubray, V. Mosseri, F. Brunin, C. Jaulerry, P. Poncet, J. Rodriguez, J. Brugère, D. Point, P. Giraud & J. Cosset (1996). Anemia is associated with lower local-regional control and survival after radiation therapy for head and neck cancer: a prospective study. *Radiology*, 201, 553–558.

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

V. T. Farewell (1986). Mixture models in survival analysis: are they worth the risk? *The Canadian Journal of Statistics*, 14, 257–262.

C. L. Faucett, N. Schenker & R. M. Elashoff (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association*, 93, 427–437.

A. E. Gelfand & A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

S. Geman & D. Geman (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

W. K. Hastings (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109.

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

N. Keiding, P. K. Andersen & J. P. Klein (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16, 215–224.

W. J. Krzanowski (1980). Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics*, 36, 493–499.

W. J. Krzanowski (1982). Mixtures of continuous and categorical variables in discriminant analysis: a hypothesis testing approach. *Biometrics*, 38, 991–1002.

A. Y. C. Kuk & C. H. Chen (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79, 531–541.

M. G. Larson & G. E. Dinse (1985). A mixture model for the regression analysis of competing risks data. *Applied Statistics*, 34, 201–211.

R. J. A. Little & M. D. Schluchter (1985). Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika*, 72, 497–512.

R. Maller & X. Zhou (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York.

W. Q. Meeker (1987). Limited failure population life tests: application to integrated circuit reliability. *Technometrics*, 29, 51–65.

X.-L. Meng (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, 9, 538–558.

X. Meng & N. Schenker (1999). Maximum likelihood estimation for linear regression models with right censored outcomes and missing predictors. *Computational Statistics and Data Analysis*, 29, 471–483.

I. Olkin & R. F. Tate (1961). Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics*, 32, 448–465.

Y. Peng, K. B. G. Dear & J. W. Denham (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine*, 17, 813–830.

D. B. Rubin (1976). Inference and missing data. *Biometrika*, 63, 581–592.

D. B. Rubin (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.

J. L. Schafer (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London.

M. D. Schluchter & K. L. Jackson (1989). Log-linear analysis of censored survival data with partially observed covariates. *Journal of the American Statistical Association*, 84, 42–52.

A. F. M. Smith & G. O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B*, 55, 53–102.

J. P. Sy & J. M. G. Taylor (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, 56, 113–122.

J. M. G. Taylor (1995). Semiparametric estimation in failure time mixture models. *Biometrics*, 51, 899–907.

H. R. Withers, L. J. Peters, J. M. G. Taylor, J. B. Owen, W. H. Morrison, T. E. Schultheiss, T. Keane, B. O'Sullivan, J. van Dyk, N. Gupta, C. C. Wang, C. U. Jones, K. P. Doppke, S. Myint, M. Thompson, J. T. Parsons, W. M. Mendenhall, S. Dische, E. G. A. Aird, J. M. Henk, M. A. M. Bidmean, V. Svoboda, Y. Chon, A. L. Hanlon, T. L. Peters & G. E. Hanks (1995). Local control of carcinoma of the tonsil by radiation therapy: an analysis of patterns of fractionation in nine institutions. *International Journal of Radiation Oncology and Biological Physics*, 33, 549–562.

K. Yamaguchi (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employment" in Japan. *Journal of the American Statistical Association*, 87, 284–292.

D. Zhuang (1999). *Survival Analysis with Long-Term Survivors and Partially Observed Covariates*. Doctoral dissertation, Department of Biostatistics, University of California, Los Angeles, CA, USA.

D. Zhuang, N. Schenker, J. M. G. Taylor, V. Mosseri & B. Dubray (2000). Analysing the effects of anæmia on local recurrence of head and neck cancer when covariate values are missing. *Statistics in Medicine*, 19, 1237–1249.

Received 9 April 1999
Accepted 13 December 2000

Meehyung CHO: meehyung_cho@merck.com
Dept. of Clinical Biostatistics and Research Data Systems
Merck Research Laboratories, Merck & Co., Inc.
P. O. Box 2000, Rahway, NJ 07065, USA

Nathaniel SCHENKER: nschenker@cdc.gov
National Center for Health Statistics, Room 915
6525 Belcrest Road, Hyattsville, MD 20782, USA

Jeremy M. G. TAYLOR: jmgt@umich.edu
Dept. of Biostatistics, University of Michigan
1420 Washington Heights, Ann Arbor, MI 48109, USA

Dongliang ZHUANG: dzhuang@scirex.com
SCIREX Corporation, Suite 450
8910 University Center Lane, San Diego, CA 92122, USA