

# Survival Analysis Using Auxiliary Variables Via Multiple Imputation, with Application to AIDS Clinical Trial Data

Cheryl L. Faucett,<sup>1</sup> Nathaniel Schenker,<sup>2</sup> and Jeremy M. G. Taylor<sup>3,\*</sup>

<sup>1</sup>Department of Biostatistics, UCLA School of Public Health,  
10833 Le Conte Avenue, Los Angeles, California 90095-1772, U.S.A.

<sup>2</sup>Office of Research and Methodology, National Center for Health Statistics,  
6525 Belcrest Road, Room 915, Hyattsville, Maryland 20782, U.S.A.

<sup>3</sup>Department of Biostatistics, School of Public Health, University of Michigan,  
1420 Washington Heights, Ann Arbor, Michigan 48109, U.S.A.

\**email*: jmgmt@umich.edu

**SUMMARY.** We develop an approach, based on multiple imputation, to using auxiliary variables to recover information from censored observations in survival analysis. We apply the approach to data from an AIDS clinical trial comparing ZDV and placebo, in which CD4 count is the time-dependent auxiliary variable. To facilitate imputation, a joint model is developed for the data, which includes a hierarchical change-point model for CD4 counts and a time-dependent proportional hazards model for the time to AIDS. Markov chain Monte Carlo methods are used to multiply impute event times for censored cases. The augmented data are then analyzed and the results combined using standard multiple-imputation techniques. A comparison of our multiple-imputation approach to simply analyzing the observed data indicates that multiple imputation leads to a small change in the estimated effect of ZDV and smaller estimated standard errors. A sensitivity analysis suggests that the qualitative findings are reproducible under a variety of imputation models. A simulation study indicates that improved efficiency over standard analyses and partial corrections for dependent censoring can result. An issue that arises with our approach, however, is whether the analysis of primary interest and the imputation model are compatible.

**KEY WORDS:** AIDS; Change-point model; Dependent censoring; Longitudinal data; Missing data; Random effects.

## 1. Introduction

### 1.1 Overview

In survival analysis, common goals are to estimate the survival distribution and the effect of time-fixed covariates on this distribution. Standard estimation methods for censored data can be used for these problems. In many medical studies, however, there is other information measured periodically on the subjects, and such measurements may be informative about the state of health of the subjects. Examples of such time-dependent markers are CD4 count and viral load in studies of HIV/AIDS. Because these markers are frequently associated with the event time, they have many possible uses, including as potential surrogate endpoints or as auxiliary variables. In this article, we will focus on their use as auxiliary variables. We will describe a method that seeks to recover information lost due to censoring by incorporating the information supplied by time-dependent markers. The goal is to make better and more efficient inferences about clinical endpoints of interest.

In a clinical trial, incorporating auxiliary variables has

potential for significantly improving the efficiency of the comparison between the arms of the trial and for reducing bias due to dependent censoring. Phase III trials typically use a meaningful clinical endpoint as the primary measure of efficacy, even in situations where most subjects will take a long time to progress to the endpoint. This creates logistical problems and expense because a large number of subjects and/or a long follow-up time may be needed. Statistical methodology that can reduce the amount of follow-up time or the number of subjects and still retain power and validity would have major implications for the conduct of clinical trials.

A number of authors (e.g., Robins and Rotnitzky, 1992; Finkelstein and Schoenfeld, 1994; Fleming et al., 1994; Gray, 1994; Malani, 1995; Murray and Tsiatis, 1996) have developed various methods for using extra auxiliary variable information to improve survival estimates, such as weighted Kaplan–Meier estimation and use of estimated likelihoods. In this article, we take a more direct approach in which we regard the event times for the censored observations as missing

data. We multiply impute values for these missing data and then analyze the augmented data, a technique that has been investigated before in different contexts (Wei and Tanner, 1990, 1991; Schenker and Taylor, 1996). If the auxiliary variables used to build the imputation model are predictive of the event times, then by incorporating information from the observed auxiliary variables, analyses of the multiply-imputed data should be more efficient than standard analyses of the data without imputation. Moreover, if censoring is related to the auxiliary variables, then imputation should lead to a partial adjustment for dependent censoring in analyses of the multiply-imputed data; such an adjustment would not be made in standard analyses that do not incorporate the auxiliary variables.

The multiple-imputation procedure we use requires a joint model for the auxiliary CD4 markers and the survival endpoints. A number of different joint models have been suggested (Clayton, 1991; Pawitan and Self, 1993; De Gruttola and Tu, 1994; Tsiatis, De Gruttola, and Wulfsohn, 1995; Faucett and Thomas, 1996; Hogan and Laird, 1997; Wulfsohn and Tsiatis, 1997). We use a similar, but more general, random effects model than used by Faucett and Thomas (1996) for the marker data, together with a proportional hazards model for the failure times. As is discussed in Section 2.5, an issue to be considered is whether such an imputation model is compatible with the analysis of primary interest.

### 1.2 AIDS Clinical Trial

Our techniques will be applied to data from the ACTG-019 clinical trial (Volberding et al., 1990). This was a double-blind randomized trial examining the efficacy of zidovudine (ZDV) in prolonging survival and delaying the onset of AIDS and advanced AIDS-related complex. Asymptomatic HIV-positive adults with CD4 counts of less than 500 were randomized to one of three treatments: placebo (428 subjects), ZDV 500 mg per day (453 subjects), or ZDV 1500 mg per day (456 subjects). For each subject, several baseline characteristics (e.g., age) were recorded, and CD4 measurements were scheduled to be obtained at 0, 8, 16, 32, 48, 64, and 80 weeks postrandomization (the median number of CD4 measurements was four). Times of diagnosis with AIDS-related complex, AIDS, or death were recorded.

We focus on time to AIDS as the primary outcome of interest. The number of subjects that were diagnosed with AIDS was 33 (7.7%) in the placebo arm and 25 (2.8%) in the treated arms combined. The median follow-up time was 50 weeks and the maximum was 108 weeks. Since preliminary analyses showed no significant differences in AIDS-free survival between the low-dose and high-dose ZDV arms, we combine these groups. Previous analyses of the CD4 count data from this trial (Bycott and Taylor, 1998) have suggested that ZDV causes an initial increase in CD4 counts for about 8 to 16 weeks, after which it tends to decline, although there is considerable variation between people.

The primary analysis goal is to estimate the effect of treatment on AIDS-free survival. Most subjects (96%) were censored. We will use the repeated CD4 measurements as auxiliary information to multiply impute times of AIDS development for censored subjects.

## 2. Methods and Assumptions

### 2.1 General Approach

For the  $i$ th subject,  $i = 1, \dots, N$ , let  $T_i^0$  be the true failure time and  $C_i$  be the censoring time so that we observe  $T_i = \min(T_i^0, C_i)$  and indicator  $D_i$  that is equal to one if the subject failed at  $T_i$  and is equal to zero if the subject was censored at  $T_i$ . Also, for subject  $i$ , let  $\mathbf{y}_i = \{y_{i1}, \dots, y_{in_i}\}$  denote the vector of  $n_i$  observations of the longitudinal auxiliary variable,  $\mathbf{x}_i$  denote fully observed fixed covariates, and  $\mathbf{z}_i = \{z_i(t), 0 \leq t \leq T_i\}$  denote the complete history of the expected value of the longitudinal auxiliary variable up to time  $T_i$ .

Our primary goal is to draw inferences about the distribution of  $T^0$  given  $\mathbf{x}^0$ , i.e.,  $[T^0 | \mathbf{x}^0]$ , where  $\mathbf{x}^0$  is a subset of  $\mathbf{x}$  and the subscript  $i$  is omitted. We call  $[T^0 | \mathbf{x}^0]$  the primary analysis model. The standard approach would be to estimate the parameters of this distribution using the observed values of  $(T, D, \mathbf{x}^0)$ . However, estimates from such an analysis would be less efficient than they would be if  $T^0$  were observed for every subject. If  $\mathbf{y}$  is predictive of  $T^0$ , incorporating  $\mathbf{y}$  into our inferences about  $[T^0 | \mathbf{x}^0]$  via multiple imputation should allow recovery of some of the information about  $T^0$  that is lost due to censoring. Moreover, incorporating  $\mathbf{y}$  via multiple imputation allows inferences about  $[T^0 | \mathbf{x}^0]$  that are valid if censoring depends on  $T^0$  through both  $\mathbf{x}$  and  $\mathbf{y}$ . Standard analyses based on  $T$ ,  $D$ , and  $\mathbf{x}^0$  require the more restrictive assumption that censoring can depend on  $T^0$  often only through  $\mathbf{x}^0$ . The Appendix gives specific assumptions under which the multiple-imputation approach is valid.

We incorporate  $\mathbf{y}$  into our analysis by imputing  $T_i^0$  for subjects with  $D_i = 0$  from the posterior predictive distribution  $[T^0 | \mathbf{T}, \mathbf{D}, \mathbf{X}, \mathbf{Y}]$ , where  $\mathbf{T}^0$  denotes the vector  $T_1^0, \dots, T_N^0$  and similarly for  $\mathbf{T}$ ,  $\mathbf{D}$ ,  $\mathbf{X}$ , and  $\mathbf{Y}$ . The imputed event times are created in two steps. First, we obtain random draws from the posterior distribution  $[\Omega | \mathbf{T}, \mathbf{D}, \mathbf{Y}, \mathbf{X}]$ , where  $\Omega$  denotes the parameters of the assumed model for  $[T^0, \mathbf{y} | \mathbf{x}]$ . We refer to  $[T^0, \mathbf{y} | \mathbf{x}]$  as the joint survival/auxiliary model, and this model will be used for imputation. As direct generation of  $\Omega$  from its posterior distribution is infeasible, we accomplish this using Markov chain Monte Carlo (MCMC) methods. Second, we impute the unobserved  $T_i^0$  conditional on the current draw of  $\Omega$  and the observed data. This two-step procedure is repeated  $M$  times, resulting in  $M$  augmented data sets. For each augmented data set, selected aspects of the distribution  $[T^0 | \mathbf{x}^0]$  are estimated, and the  $M$  estimates are then combined using standard multiple-imputation procedures (Rubin and Schenker, 1986; Rubin, 1987). The following sections describe the assumed models and application of this general approach to the analysis of the AIDS clinical trial data.

### 2.2 Primary Analysis Model

For our application,  $T^0$  denotes the time of AIDS development and  $\mathbf{x}^0$  is the scalar treatment indicator, with  $\mathbf{x}^0 = 1$  for ZDV and  $\mathbf{x}^0 = 0$  for placebo. We consider two primary analysis models,  $[T^0 | \mathbf{x}^0]$ . The first is a nonparametric model, in which the quantities of primary interest are the AIDS-free survival

probabilities for the two treatment groups. The second is a Cox proportional hazards model with hazard  $\lambda(t) = \lambda_0(t) \times \exp(\gamma^0 x^0)$ , in which the quantity of primary interest is the relative hazard of AIDS for treatment versus placebo ( $\gamma^0$ ). See Section 2.4.3 for further discussion.

### 2.3 Joint Survival/Auxiliary Model

In our application, the auxiliary variables  $\mathbf{y}$  are the repeated measurements of CD4. The joint distribution,  $[T^0, \mathbf{y} \mid \mathbf{x}, \Omega]$ , can be factored into a longitudinal CD4 model,  $[\mathbf{y} \mid \mathbf{x}, \Omega]$ , and a conditional model for time to AIDS given CD4,  $[T^0 \mid \mathbf{y}, \mathbf{x}, \Omega]$ .

**2.3.1 Longitudinal CD4 model.** Let  $t_{ij}$  denote the time since randomization at which  $y_{ij}$  was observed. For each subject, we assume that CD4 level (after a transformation) follows a piecewise-linear random effects model with transition point  $\tau_i$ . Specifically,  $y_{ij} = z_i(t_{ij}) + \epsilon_{ij}$ , where  $z_i(t) = b_{0i} + b_{1i}\{t - (t - \tau_i)^+\} + b_{2i}(t - \tau_i)^+$ , and  $(t - \tau)^+$  equals zero for  $t \leq \tau$  and equals  $t - \tau$  for  $t > \tau$ . The random effects include subject-specific intercepts,  $b_{0i}$ , and pre- and posttransition point slopes,  $b_{1i}$  and  $b_{2i}$ , respectively. The vectors  $\mathbf{b}_i = (b_{0i}, b_{1i}, b_{2i})'$  are assumed to be i.i.d. multivariate normal with mean  $(\mathbf{x}'_{0i}\beta_{0s}, \mathbf{x}'_{1i}\beta_{1s}, \mathbf{x}'_{2i}\beta_{2s})'$  and covariance matrix  $\Sigma_s$ , where  $s = 1, \dots, S$  denotes a subgroup to which subject  $i$  belongs. The vectors  $\mathbf{x}_{ki}$  and  $\beta_{ks}$ ,  $k = 0, 1, 2$  (each of dimension  $p_{ks}$ ), contain, respectively, a subset of the measured covariates for subject  $i$  and corresponding vectors of fixed effects. The logarithms of the transition points  $\log_e(\tau_i)$  are assumed to be i.i.d.  $N(\mathbf{x}'_{\tau i}\beta_{\tau s}, \sigma_{\tau s}^2)$ , where  $\mathbf{x}_{\tau i}$  also contains measured covariates for subject  $i$  and  $\beta_{\tau s}$  is the corresponding vector of fixed effects for subgroup  $s$ . The  $\epsilon_{ij}$  are assumed to be i.i.d.  $N(0, \sigma_{\epsilon s}^2)$ .

**2.3.2 AIDS hazard model.** We assume that failure times follow a proportional hazards model that is a function of the expected CD4 level  $z_i(t)$  and covariates  $\mathbf{x}_i$ , i.e.,  $\lambda\{t \mid \mathbf{y}_i, \mathbf{x}_i, z_i(t)\} = \lambda_0(t) \exp\{\phi z_i(t) + \gamma' \mathbf{x}_i\}$ . Note that, conditional on  $z_i(t)$ , the hazard is independent of  $\mathbf{y}_i$ . The corresponding survival function is  $S(t) = \exp[-\int_0^t \lambda\{s \mid \mathbf{x}_i, z_i(s)\} ds]$ . The baseline hazard function  $\lambda_0(t)$  is assumed to be a step function on a predefined set of time intervals  $L_1, \dots, L_K$ , i.e.,  $\lambda_0(t) = \lambda_k$ ,  $t \in L_k$ ,  $k = 1, \dots, K$ .

**2.3.3 Variations of the joint model.** In our application to the ACTG-019 data, we will impute from several variations of the joint survival/auxiliary model in order to examine the sensitivity of the results to model specification.

In the first model (model M1), a single longitudinal model is assumed ( $S = 1$ ) for  $y_{ij} = (\text{CD4})^{1/2} - 18.0$ . A common intercept is assumed for the two treatment groups ( $\mathbf{x}_{0i} = 1$  for all subjects), but treatment effects are included for both slopes ( $\mathbf{x}_{1i} = \mathbf{x}_{2i} = (1, 0)'$  for subjects given placebo,  $\mathbf{x}_{1i} = \mathbf{x}_{2i} = (1, 1)'$  for treated subjects). We constrain each subject to have a common transition point,  $\tau_i = 56$  days. Thus, we set  $p_{rs} = 1$ ,  $x_{ri} = 1$ ,  $\beta_{rs} = 4.03$ , and  $\sigma_{\tau s}^2 = 0$ . The expected CD4 level at time  $t$ ,  $z_i(t)$ , and the treatment assignment indicator are included in the AIDS hazard model. We use  $K = 19$  baseline hazard steps, each of width 40 days.

Each of models M2 through M6 can be described in terms of how it differs from model M1. In model M2, we set  $\tau_i = 112$  days instead of  $\tau_i = 56$  days. In model M3, subject-specific transition points are allowed, i.e.,  $\sigma_{\tau s}^2$  is not fixed at 0. For model M4,  $y_{ij} = (\text{CD4})^{1/3} - 6.82$ . In model M5, CD4 is

modeled separately for the placebo ( $s = 1$ ) and treated ( $s = 2$ ) groups. Within each treatment arm, a single population intercept and single pre- and posttransition point slopes are specified, i.e.,  $p_{ks} = 1$  and  $\mathbf{x}_{ki} = 1$  for  $s = 1, 2$ ,  $k = 0, 1, 2$ , and  $i = 1, \dots, N$ . Model M6 is the same as model M1 except that age is also included as a covariate in the hazard model. Finally, the treatment-only imputation model includes only a treatment indicator in the hazard model.

### 2.4 Multiple Imputation Procedure

**2.4.1 Sampling joint survival/auxiliary model parameters.** In order to impute  $T^0$ , we first generate samples from the posterior distribution  $[\Omega \mid \mathbf{T}, \mathbf{D}, \mathbf{Y}, \mathbf{X}]$ . The posterior density of the unknown parameters and random effects in the joint model is proportional to a product of the likelihoods based on the longitudinal CD4 and AIDS hazard models and the densities for the prior distributions. Weakly informative prior distributions are adopted for the population parameters. These include uniform priors for  $\beta_{ks}$ ,  $\beta_{\tau s}$ ,  $\phi$ , and  $\gamma$ , inverse-Wishart priors for  $\Sigma_s$ , inverse gamma priors for  $\sigma_{\epsilon s}^2$  and  $\sigma_{\tau s}^2$ , and gamma priors for  $\lambda_k$ . The prior for  $\beta_{\tau s}$  constrains  $\tau$  to be between 50 to 300 days since we have limited information regarding  $\tau$  outside that range and we expect an early ZDV effect.

Since sampling directly from the joint distribution is infeasible, we rely on MCMC methods (Geman and Geman, 1984; Li, 1988; Gelfand and Smith, 1990). After assignment of initial values to all parameters, each unspecified parameter is sampled in turn from its full conditional distribution given the observed data and the current values of the other parameters. Specifically, each  $\lambda_k$  is sampled from a gamma distribution, and  $\phi$  and each component of  $\gamma$  are sampled univariately using adaptive rejection sampling (Gilks, 1992). The random effects  $\mathbf{b}_i$  and  $\tau_i$  are sampled jointly for each subject using a Metropolis–Hastings step (Gilks, Richardson, and Spiegelhalter, 1995), and the  $\beta_{ks}$  and  $\beta_{\tau s}$  are sampled from truncated normal distributions. The  $\Sigma_s$  matrices are sampled from inverse Wishart distributions, and  $\sigma_{\tau s}^2$  and  $\sigma_{\epsilon s}^2$  are sampled from inverse gamma distributions. For each imputation model, 500 iterations were eliminated for convergence and 10,000 additional iterations were generated.

**2.4.2 Imputation of event times.** Given a set of sampled values for all parameters, a new event time and disease status for each censored subject is imputed. A random value  $U$  is drawn from a uniform distribution on  $(0, 1)$  and then the equation  $S\{t \mid t > T_i, z_i(t), \mathbf{x}_i\} = U$  is solved for  $t$ . If the resulting value of  $t$  exceeds a prespecified maximum allowable time,  $t_{\max}$ , then the subject is considered censored at time  $t_{\max}$ . For the AIDS clinical trial data, we set  $t_{\max} = 720$  days, the end of the hazard interval that included the last observed event time. If  $t \leq t_{\max}$ , then the imputed event time is set to  $t$  and  $D_i$  is set to one. Imputations were produced after every 100th iteration of the MCMC procedure, resulting in  $M = 100$  augmented data sets. Other values for  $t_{\max}$  were also explored for model M1.

By limiting the range of the imputed failure times, we avoid extrapolation beyond the range of the observed data and thus reduce the effects of the imputation model assumptions on some multiple-imputation analyses. The influence of the assumed imputation model can also be reduced by this limit since only small changes in survival time can be imputed for

censored subjects with long follow-up times. Because of this limit, however, if there is CD4-dependent censoring, then the imputation approach can only partially adjust for this dependency in some analyses (see Remark C of the Appendix). Use of a good imputation model should improve estimates compared with standard approaches that make no adjustment for dependent censoring, however.

**2.4.3 Analysis of augmented data.** Each of the  $M$  augmented data sets is analyzed using two methods of analysis, and the resulting analyses are combined using standard methods (Rubin and Schenker, 1986; Rubin, 1987). First, a Kaplan–Meier estimate of the probability of remaining AIDS-free at a given time  $t$  is calculated for each treatment group. A variance estimate for this estimated probability at each observed (not imputed) time of AIDS development is obtained by applying Greenwood’s formula. Second, Cox regression is used to estimate the log hazard ratio ( $\gamma^0$ ) of the proportional hazards model of Section 2.2. Here, in contrast with the joint survival/auxiliary model that we use to produce imputed AIDS times, we make no assumptions regarding the form of the baseline hazard.

For comparison, we perform Kaplan–Meier and Cox regression analyses of the original data  $(\mathbf{T}, \mathbf{D}, \mathbf{X})$  to estimate components of the distribution  $[T^0 \mid \mathbf{x}^0]$ . We call this the partially observed analysis.

## 2.5 Incompatibility Between Models

A key issue in our approach is whether the primary analysis model used for analyzing the augmented data and the joint survival/auxiliary model used in creating the imputations are incompatible. This is closely related to the issue of uncongeniality discussed in Meng (1994) and Rubin (1996). We briefly discuss two types of situations that are relevant to our problem.

The first type of incompatibility occurs when one model is correct but the other is incorrect. In such a situation, inferences based on multiple imputation are likely to be incorrect as well. For example, the proportional hazards model for  $[T^0 \mid \mathbf{x}^0]$  described in Section 2.2 is incompatible with the AIDS hazard model described in Section 2.3.2 because proportionality of hazards when a set of predictors is included generally implies nonproportionality of hazards when only a subset of the predictors is included. This would explain the apparent biases of the estimates of the treatment effect  $\gamma^0$ , both with and without multiple imputation, in the simulation results of Section 4.2.2 when the data are generated under our assumed joint survival/auxiliary model with independent censoring but then analyzed using a proportional hazards primary analysis model. The simulations also suggest, however, that inferences with multiple imputation are better than inferences without multiple imputation.

The second type of incompatibility occurs when both models are correct but extra information about the quantity of primary interest that is used in creating imputations is then not used in analyzing the augmented data. We suspect that this represents a type of uncongeniality, discussed in Meng (1994) and Rubin (1996), under which inferences with multiple imputation tend to be conservative but more efficient than inferences without multiple imputation. This is consistent with the simulation results of Section 4.2.1 when Kaplan–Meier estimation is used as the primary analysis. The extra information in this case could be the model structure

that is present in the creation of imputations but not present in the Kaplan–Meier analysis. See also Remark B of the Appendix.

For our problem, the first type of incompatibility, in which one of the models is incorrect, is a much greater concern than the second type. Thus, in practice, careful consideration and checking of models should be carried out so that neither model will be obviously incorrect. Moreover, sensitivity analysis could be used to assess the magnitude of possible biases.

## 3. Analyses of AIDS Clinical Trial Data

### 3.1 Kaplan–Meier Estimation

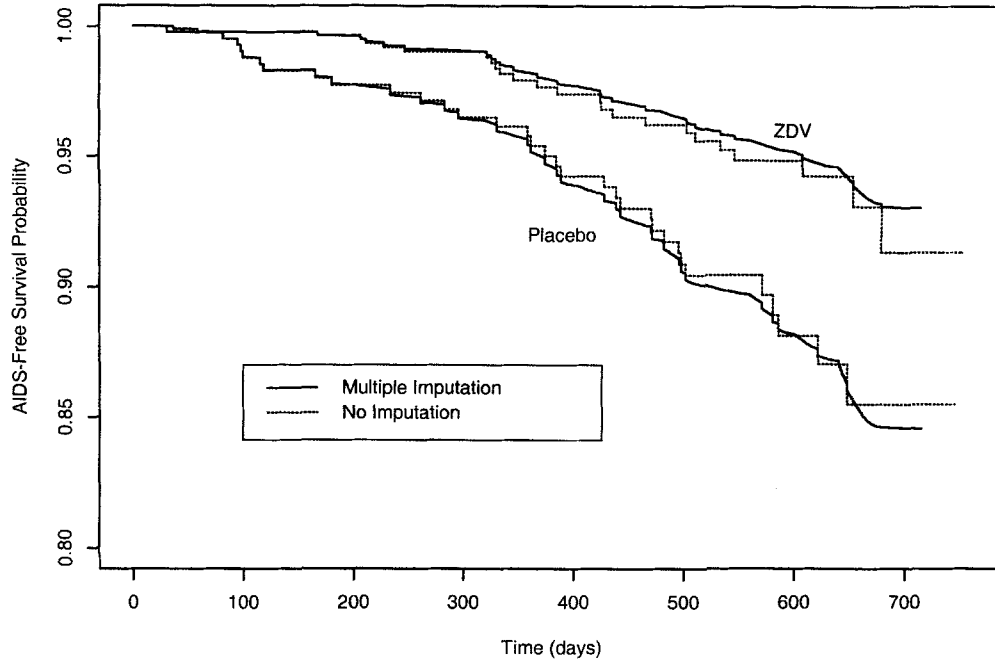
Figure 1 displays the estimated Kaplan–Meier curves from the partially observed (PO) analysis and from the multiple-imputation (MI) analysis under model M1. For the placebo group, the two analyses produce nearly identical estimated probabilities. On the other hand, for the treated group, the MI method yields larger estimated survival probabilities toward the end of the observation time. A possible interpretation is that the censored subjects in the treated group tended to be healthier and, if they could have been followed longer, would have developed AIDS at a lower rate than those that were not censored. Thus, the MI analysis could be providing an adjustment for dependent censoring.

Figure 2 displays the ratio of the estimated variance of the survival probability estimate from the PO analysis to that from the MI analysis based on model M1. The larger ratios occur later in the observation time since there is more censoring at later times in the original data set and thus more imputation of events at those times.

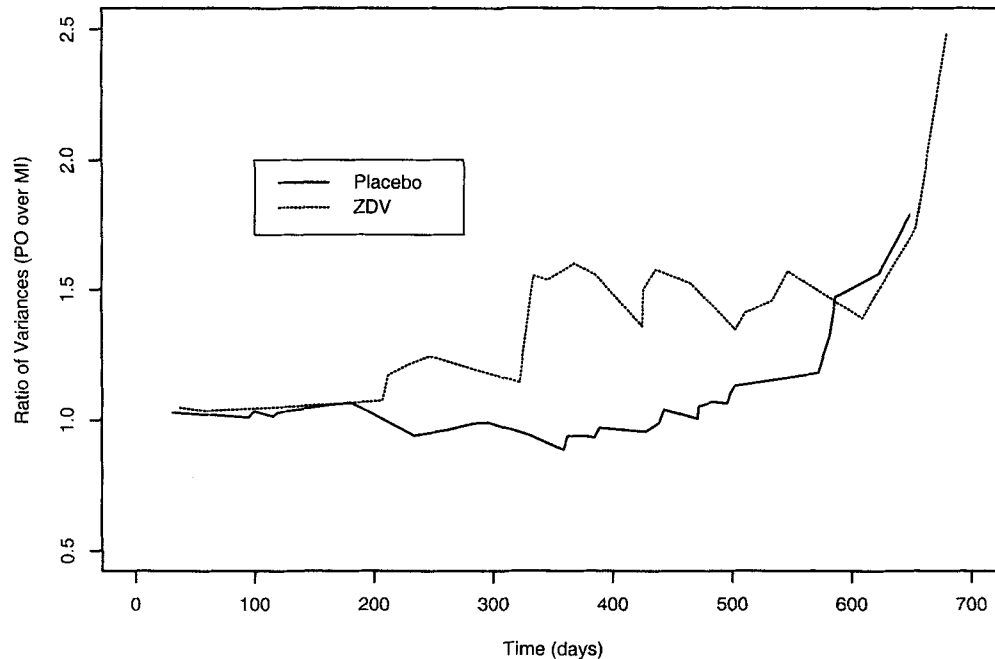
Table 1 provides a comparison of Kaplan–Meier estimates and estimated standard errors at two points in time for the PO method and the MI method under the various imputation models. The survival estimates from MI models M1–M6 are similar to each other but slightly higher than those obtained from the PO analysis for the treated group. Multiple imputation also leads to a reduction in the estimated standard errors compared with the PO analysis. We suspect that the smaller standard errors under the treatment only (TO) imputation model when compared with the PO analysis are due to use of a parametric hazard specification in the imputation procedure and possibly the use of weak informative priors, both of which could add information to the nonparametric Kaplan–Meier analysis (see Section 2.5).

### 3.2 Cox Regression

Table 2 displays the estimated treatment effect  $\hat{\gamma}^0$  and estimated standard error based on the PO analysis and based on the MI approach under the various imputation models. There is little difference between the PO and TO estimates, as expected, since no auxiliary information is utilized and since the primary analysis is more parametric than the Kaplan–Meier analysis of Section 3.1. When we add CD4 to the imputation model (models M1–M6), all estimated standard errors are smaller than in the PO analysis. The ratio of the estimated variance obtained from the PO analysis to that obtained from the MI analyses ranges from 108 to 127%. The  $\gamma^0$  estimates are larger for models M1–M6, which is consistent with the results for Kaplan–Meier estimation in Section 3.1. Model M6, which uses both CD4 and age as auxiliary variables, yields the smallest estimated standard error.



**Figure 1.** Comparison of Kaplan–Meier curves based on the partially observed data (no imputation) and based on multiple imputation under model M1.



**Figure 2.** Ratio of estimated variances of survival probability estimates based on the partially observed data (no imputation) to those based on multiple imputation under model M1.

Table 3 displays the results for model M1 when imputation of an event time is limited to a specific number of days (rather than the maximum limit of 720 days). In comparison with the PO analysis, the MI method results in larger estimates of treatment effect and smaller estimated standard errors regardless of the limit placed on the event times.

#### 4. Simulation Study

We performed a simulation study to evaluate the performance of our method under a variety of parameter combinations. In our base simulation, data were generated to approximate the structure of the ZDV clinical trial, although we reduced the sample size and number of imputations, simplified the model,

**Table 1**  
Kaplan–Meier estimates of AIDS-free survival probabilities and estimated standard errors based on partially observed data (no imputation) and multiply-imputed data

Model	Treatment group				Treatment group			
	Placebo		Treated		Placebo		Treated	
	$\hat{S}(450)^a$	$(\widehat{SE}_{400-500})^b$	$\hat{S}(450)$	$(\widehat{SE}_{400-500})$	$\hat{S}(550)$	$(\widehat{SE}_{500-600})$	$\hat{S}(550)$	$(\widehat{SE}_{500-600})$
Partially observed	0.930	(0.01542)	0.965	(0.00842)	0.905	(0.01968)	0.949	(0.01040)
Treatment only	0.930	(0.01519)	0.969	(0.00712)	0.907	(0.01848)	0.954	(0.00894)
M1	0.927	(0.01516)	0.971	(0.00690)	0.903	(0.01744)	0.957	(0.00865)
M2	0.928	(0.01521)	0.971	(0.00711)	0.905	(0.01826)	0.957	(0.00846)
M3	0.928	(0.01515)	0.971	(0.00716)	0.905	(0.01848)	0.957	(0.00869)
M4	0.926	(0.01582)	0.973	(0.00665)	0.903	(0.01923)	0.958	(0.00841)
M5	0.928	(0.01541)	0.971	(0.00704)	0.906	(0.01840)	0.958	(0.00876)
M6	0.927	(0.01519)	0.971	(0.00686)	0.904	(0.01847)	0.957	(0.00825)

<sup>a</sup> Kaplan–Meier estimated probability of remaining AIDS-free at day 450.

<sup>b</sup> Greenwood’s formula averaged over days 400–500 in the complete-data standard error estimate.

and increased effect sizes. Each additional simulation varied one of the base model parameters. We present results for both an estimated survival probability and the estimate of  $\gamma^0$  from the model  $\lambda(t) = \lambda_0(t) \exp(\gamma^0 x^0)$ .

#### 4.1 Data Generation

Data were simulated for 400 subjects, with 200 receiving placebo and 200 receiving treatment. The parameter values given next are those used in the base simulation. A normal linear model was used for  $y_{ij}$ , i.e.,  $y_{ij} = z_i(t_{ij}) + \epsilon_{ij}$ , where  $\epsilon_{ij} \sim N(0, 2.5^2)$  and  $z_i(t) = b_{0i} + b_{1i}t$ , with  $b_{0i} \sim N(0, 9)$  and  $b_{1i} \sim N(\mu_{TX}, 0.002^2)$ . The value of  $\mu_{TX}$  was  $-0.008$  for the placebo group and  $0.0$  for the treated group. Observations of  $y_{ij}$  were generated at time 0 and every 3 months thereafter, for a total study length of 2 years and a maximum of 8 measurements. The time of AIDS development ( $T_i^0$ ) was generated from the model  $\lambda(t) = \lambda_0(t) \exp\{\phi z_i(t) + \gamma x_i\}$ , where  $\lambda_0(t) = 0.0008$ ,  $\phi = -0.5$ ,  $\gamma = -0.5$ , and  $x_i$  is the treatment indicator. A random censoring time ( $C_i$ ) was generated based on the model  $\lambda_c(t) = \exp\{\psi_0 + \psi_1 z_i(t)\}$ , with  $\psi_0 = -5.0$  and  $\psi_1 = 0.0$ .

One hundred data sets were generated and analyzed to obtain survival probability estimates at 1 year for each treatment group and estimates of the marginal treatment effect  $\gamma^0$  using three methods of analysis. First, for the fully observed (FO) analysis, which we treat as the gold standard, we applied the primary analysis (Kaplan–Meier estimation or Cox regression) to each data set without any random censoring so that  $T_i$  was equal to  $\min(T_i^0, 2 \text{ years})$ . Second, for the usual partially observed (PO) analysis, we applied the primary analysis to each data set with random censoring so that  $T_i$  was equal to  $\min(T_i^0, C_i, 2 \text{ years})$ . Finally, the multiple-imputation (MI) analysis was based on the same data as the PO analysis, but it used the auxiliary variables to impute event times past  $T_i$  but prior to 2 years. Fifty augmented data sets ( $M = 50$ ) were created, the primary analysis was applied to each data set, and the results of the 50 analyses were combined.

We computed the mean and standard deviation of estimates across the 100 data replicates. We also computed the average over the 100 replicates of the squared difference

of the PO and MI estimates, respectively, from the FO estimates. This latter measure provides a single summary of the closeness of the PO and MI estimates to the FO estimates. We computed the squared-difference ratio (SDR), with values above 1.0 indicating improved estimation by the MI method.

#### 4.2 Simulation Results

**4.2.1 Kaplan–Meier estimation.** Over a range of parameter combinations, except in the case of dependent censoring, there was little bias associated with the MI method and the MI estimates were considerably less variable than the PO estimates. This was true even when there was no effect of the auxiliary variable on disease risk ( $\phi = 0.0$ , model 4), which is consistent with the results for the TO imputation model in Section 3.1. The reason for the gain in efficiency is that extra information, specifically the assumptions of proportional hazards and a constant baseline hazard as well as the prior distributions, is used in the creation of the imputes but not in the primary analysis (also see Section 2.5). The MI method was better at reproducing the FO estimates than was the PO analysis, with the SDR ranging from 1.8 to 4.7.

**Table 2**  
Estimated treatment effect from Cox regression based on partially observed data (no imputation) and multiply-imputed data

Model	Estimate $\hat{\gamma}^0$	Estimated standard error	t-Statistic <sup>a</sup>	p-Value
Partially observed	-0.798	0.266	3.00	0.0027
Treatment only	-0.799	0.260	3.07	0.0023
M1	-0.861	0.250	3.44	0.0006
M2	-0.832	0.249	3.34	0.0009
M3	-0.812	0.256	3.17	0.0017
M4	-0.906	0.253	3.58	0.0004
M5	-0.823	0.250	3.29	0.0011
M6	-0.842	0.236	3.56	0.0004

<sup>a</sup> z-Statistic for partially observed analysis.

**Table 3**  
*Estimated treatment effect from Cox regression based on partially observed data (no imputation) and on multiple imputation under model M1, with limited imputation times*

Limit	Estimate $\hat{\gamma}^0$	Estimated standard error	<i>t</i> -Statistic	<i>p</i> -Value	Percentage censored	
					Treated	Placebo
Partially observed	-0.798	0.266	3.00 <sup>a</sup>	0.0027	97.2	92.3
100 days	-0.817	0.263	3.10	0.0019	96.4	90.3
200 days	-0.853	0.262	3.25	0.0012	95.6	88.6
300 days	-0.870	0.264	3.30	0.0010	94.9	87.2
400 days	-0.869	0.258	3.37	0.0008	94.0	86.0
500 days	-0.860	0.252	3.41	0.0007	93.4	85.2
600 days	-0.862	0.251	3.44	0.0006	93.1	84.7
700 days	-0.861	0.250	3.44	0.0006	93.1	84.6
No limit <sup>b</sup>	-0.861	0.250	3.44	0.0006	93.1	84.6

<sup>a</sup> *z*-Statistic for partially observed analysis.

<sup>b</sup> Up to 720 days posttreatment.

With the introduction of censoring that depends on the auxiliary variable (models 11 and 12), the mean of the MI estimate was closer than the PO mean to the mean FO estimate, although the MI estimate was still biased. In models 11 and 12, the assumptions regarding independent censoring required for the validity of the PO method are violated since censoring depends on  $\mathbf{z}_i$ . To the extent that this dependence is explained by  $\mathbf{y}_i$ , the MI method corrects for it (see Assumption 2 in the Appendix).

For models 1–10, the proportion of times the MI estimate plus or minus 1.96 times its estimated standard error contained the mean FO estimate ranged from 95 to 99% (data not shown) and always exceeded the coverage rate for the PO method, which ranged from 77 to 98%. The estimated SEs were on average 16% higher than the SDs of the point estimates for the MI method. This provides some evidence that our imputation model did not produce artificially small estimates of the standard errors (also see Section 2.5 and Remark B of the Appendix).

The results for the placebo group are in general similar to those for the treated group, although for the MI method, there was a slight upward bias in the estimates for some of the models and the range of the SDR was 1.7–6.9. The results for survival probability estimates at 1.5 years were also calculated and were similar to those at 1 year. Differences included slightly more bias in the estimates for the placebo group for the MI method under some models and greater efficiency gains for the MI method compared with the PO method.

**4.2.2 Cox regression.** For estimating the marginal treatment effect  $\gamma^0$  over a range of parameter combinations, the MI method produced estimates that were less variable and closer to the FO method estimates than were the standard PO estimates, as shown in Table 4. Provided there was some effect of the auxiliary variable on disease risk, the MI method was better at reproducing the FO estimates than was the PO analysis, with the SDR ranging from 1.3 to 3.5. When there was no effect of the auxiliary variable on disease risk ( $\phi = 0.0$ , model 4), the PO and MI methods both provided approximately unbiased estimation of the mean FO estimate and there was no advantage to using the MI method compared

with the PO method, which is consistent with the results for the TO imputation method in Section 3.2.

The PO and MI methods were generally biased at estimating the FO marginal treatment effect, even under independent censoring. This was due to the omission of the auxiliary variable from the primary analysis model, which results in the proportional hazards assumption in the marginal model (i.e., the primary analysis model) not strictly holding (see Section 2.5).

Finally, the proportion of times the MI estimate plus or minus 1.96 times its estimated standard error contained the mean FO estimate ranged from 70 to 99% across the 12 simulation conditions (data not shown). The corresponding range for the PO method was 35–98%, with the PO coverage rate never exceeding the MI coverage rate. The estimated SEs were on average 21% higher than the SDs of the point estimates for the MI method. This again provides some evidence that our imputation model did not produce artificially small estimates of the standard errors.

## 5. Discussion

The simulation study shows that the use of the MI method can lead to improved performance of estimators. The MI point estimates were generally less variable and closer to those produced by the FO method than were the estimates produced by the PO method. The MI approach did not eliminate all bias, however. In the case of dependent censoring, this may be due in part to the fact that the MI approach only adjusts for censoring related to the observed values  $\mathbf{y}$  and not the latent values  $\mathbf{z}$ . In the case of the Cox proportional hazards primary analysis model, it may be due in part to a deficiency in the fit of the primary analysis model (see Section 2.5).

In the ACTG-019 trial, there was a large amount of censoring, some of which was administrative censoring due to termination of the study and some of which was loss to follow-up. The reasons for subjects dropping out of the study may have been related to their health, which could lead to biased estimates of survival probabilities due to dependent censoring. Utilizing the auxiliary CD4 variable, we found larger effect estimates and lower estimated standard errors than produced by a standard analysis of the observed data. The magnitudes of

**Table 4**  
*Simulation results: survival probability estimates at 1 year for treated subjects and marginal treatment effect parameter ( $\gamma^0$ ) estimates*

	Simulation model <sup>a</sup>					Fully observed		Partially observed		Multiple imputation		Squared difference ratio <sup>b</sup>
	$\gamma$	$\phi$	$\mu_0$	$\psi_0$	$\psi_1$	Mean	SD	Mean	SD	Mean	SD	
<b>Survival Probability Estimates at One Year for Treated Subjects</b>												
Base model												
1.	-0.5	-0.50	-0.008	-5.0	0.0	0.74	0.029	0.74	0.058	0.74	0.039	2.3
Vary treatment effect on disease												
2.	-1.0					0.81	0.026	0.82	0.052	0.81	0.036	2.4
3.	0.0					0.65	0.030	0.65	0.070	0.65	0.045	2.6
Vary auxiliary-variable effect on disease												
4.		0.00				0.84	0.024	0.83	0.056	0.84	0.036	2.8
5.		-0.25				0.81	0.024	0.81	0.053	0.80	0.041	1.9
6.		-0.75				0.68	0.030	0.67	0.062	0.68	0.040	3.1
Vary difference in $\mu$ between treatments												
7.			-0.012			0.74	0.029	0.74	0.058	0.73	0.037	2.5
8.			-0.004			0.74	0.029	0.74	0.058	0.74	0.039	2.4
Vary amount of censoring												
9.				-5.5		0.74	0.029	0.73	0.042	0.74	0.034	1.8
10.				-4.5		0.74	0.029	0.75	0.111	0.73	0.051	4.7
Add dependent censoring												
11.					-0.5	0.74	0.029	0.90	0.030	0.84	0.034	2.3
12.					-1.0	0.74	0.029	0.93	0.024	0.87	0.037	1.9
<b>Marginal Treatment Effect Parameter (<math>\gamma^0</math>)</b>												
1.	-0.5	-0.50	-0.008	-5.0	0.0	-1.12	0.12	-0.72	0.24	-0.87	0.21	2.1
2.	-1.0					-1.49	0.13	-1.15	0.26	-1.19	0.22	1.3
3.	0.0					-0.78	0.11	-0.32	0.22	-0.57	0.19	3.5
4.		0.00				-0.49	0.16	-0.51	0.31	-0.47	0.33	0.9
5.		-0.25				-1.03	0.14	-0.71	0.28	-0.87	0.26	2.1
6.		-0.75				-1.10	0.12	-0.62	0.21	-0.81	0.19	2.4
7.			-0.012			-1.47	0.12	-0.90	0.24	-1.11	0.21	2.3
8.			-0.004			-0.73	0.13	-0.55	0.25	-0.58	0.22	1.4
9.				-5.5		-1.12	0.12	-0.81	0.21	-0.92	0.19	1.9
10.				-4.5		-1.12	0.12	-0.64	0.29	-0.79	0.26	1.8
11.					-0.5	-1.12	0.12	-0.81	0.34	-1.05	0.28	2.7
12.					-1.0	-1.12	0.12	-0.92	0.35	-1.08	0.32	1.8

<sup>a</sup> Parameters are set to the base-model values unless otherwise noted. Remaining model parameter values are given in Section 4.1.

<sup>b</sup> (Partially observed  $\text{dif}^2$ )/(multiple-imputation  $\text{dif}^2$ ), where  $\text{dif}^2$  is the average across 100 replicates of the squared difference of the estimate from the fully observed estimate.

these changes were consistent with results from other articles (e.g., Murray and Tsiatis, 1996). These results could indicate improved performance of the MI method or they could be an artifact of the assumed imputation model. We addressed the latter by comparing analyses under different imputation models. The consistency of the qualitative conclusions regardless of the model used was reassuring, but results from a single data set cannot be considered conclusive evidence that our approach adjusted for dependent censoring or produced more efficient estimates. Also, although imputation models M1–M6 are flexible, none can be assumed to be perfect, and thus to

the extent that they are not “true,” they may induce some bias in the final parameter estimates.

The use of multiple imputation to handle censored observations provides two potential sources of robustness to model misspecification in addition to that achieved by avoiding extrapolation outside the range of the data. First, if a large amount of information is missing about a subject’s failure time, then the multiply-imputed failure times are likely to display a large amount of variability, which could swamp bias due to model misspecification. Second, methods based on models other than those used for imputation, such as a Weibull regres-



sion or a log-rank comparison, could be used for the analysis of the augmented data sets. Thus, the assumptions underlying the imputation method need only affect final inferences through their effect on the imputed values for the censored survival times and not through the method chosen to analyze the augmented data. As discussed in Section 2.5, however, use of a primary analysis method that is incompatible with the imputation method can result in incorrect inferences.

Our general approach of joint modeling followed by multiple imputation has the flexibility to be extended in a number of ways. Different forms could be considered for both the joint survival/auxiliary imputation model and the primary analysis models. Other time-dependent covariates could be investigated, including categorical covariates (e.g., Faucett, Schenker, and Elashoff, 1998), and important baseline prognostic categorical or continuous variables could be added to the models, as demonstrated by our adding age in model M6. We suspect that, under the correct models, superior performance of the estimates could be achieved through the use of a more accurate and/or more frequently measured marker of disease progression.

#### ACKNOWLEDGEMENTS

This research was supported in part by NIH grant AI29196 while all of the authors were in the UCLA Department of Biostatistics. The authors thank the editor, the associate editor, and the referee for their helpful suggestions.

#### RÉSUMÉ

Nous avons développé une approche fondée sur le principe de l'imputation multiple pour utiliser des variables auxiliaires afin de reconstituer l'information des données censurées dans les analyses de survie. Nous appliquerons cette approche à des données issues d'un essai clinique dans le SIDA comparant ZDV et placebo, pour lequel la numération des CD4 est la variable auxiliaire dépendant du temps. Pour faciliter l'imputation, un modèle joint est développé pour les données, qui inclut un modèle hiérarchique de changement d'état pour les CD4 et un modèle des taux proportionnels pour les délais de survenue du SIDA. La méthode de Monte Carlo pour des chaînes de Markov est utilisée pour une imputation multiple des temps d'événements aux observations censurées. La série de données ainsi étoffée est ensuite analysée et les résultats combinés en utilisant des techniques d'imputation multiple standard. Une comparaison de notre approche après imputation multiple et de la simple analyse des données observées indique que l'imputation multiple conduit à une faible modification de la mesure de l'effet du ZDV et à une estimation plus faible des écarts type. Une analyse de sensibilité suggère que ces résultats sont stables pour plusieurs modèles d'imputation. Une étude de simulation montre qu'il peut en résulter une amélioration de l'efficacité de la procédure par rapport à une analyse classique et une correction partielle en cas de censure dépendante. Un problème est néanmoins apparu avec notre approche qui est de savoir si l'analyse de premier intérêt et le modèle d'imputation sont compatibles.

#### REFERENCES

- Bycott, P. W. and Taylor, J. M. G. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statistics in Medicine* **17**, 2061–2077.
- Clayton, D. G. (1991). Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In *Statistical Models for Longitudinal Studies of Health*, J. H. Dwyer, P. Lippert, M. Feinleib, and H. Hoffmeister (eds), 301–331. New York: Oxford University Press.
- De Gruttola, V. and Tu, X. M. (1994). Modelling progression of CD4 lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.
- Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine* **15**, 1663–1685.
- Faucett, C. L., Schenker, N., and Elashoff, R. (1998). Analysis of censored survival data with intermittently observed binary time-dependent covariates. *Journal of the American Statistical Association* **93**, 427–437.
- Finkelstein, D. M. and Schoenfeld, D. A. (1994). Analysing survival in the presence of an auxiliary variable. *Statistics in Medicine* **13**, 1747–1754.
- Fleming, T. R., Prentice, R. L., Pepe, M. S., and Glidden, D. (1994). Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. *Statistics in Medicine* **13**, 955–968.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Gilks, W. R. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics*, Volume 4, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), 1–9. Oxford: Oxford University Press.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1995). Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (eds), 1–19. London: Chapman and Hall.
- Gray, R. J. (1994). A kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika* **81**, 527–540.
- Hogan, J. W. and Laird, N. M. (1997). Mixture models for the joint distribution or repeated measures and event times. *Statistics in Medicine* **16**, 239–257.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Li, K. H. (1988). Imputation using Markov chains. *Journal of Statistical Computation and Simulation* **30**, 57–79.
- Malani, H. M. (1995). Modification of the re-distribution to the right algorithm using disease markers. *Biometrika* **82**, 515–526.
- Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* **9**, 538–557.
- Murray, S. and Tsiatis, A. A. (1996). Non-parametric survival estimation using prognostic longitudinal covariates. *Biometrics* **52**, 137–151.

- Pawitan, Y. and Self, S. (1993). Modeling disease marker processes in AIDS. *Journal of the American Statistical Association* **88**, 719–726.
- Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology: Methodological Issues*, N. P. Jewell, K. Dietz, and V. T. Farewell (eds), 297–331. Boston: Birkhauser.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–590.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91**, 473–489.
- Rubin, D. B. and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association* **81**, 366–374.
- Schenker, N. and Taylor, J. M. G. (1996). Partially parametric techniques for multiple imputation. *Computational Statistics and Data Analysis* **22**, 425–446.
- Tsiatis, A. A., De Gruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90**, 27–37.
- Volberding, P. A., Lagakos, S. W., Koch, M. A., et al. (1990). Zidovudine in asymptomatic human immunodeficiency virus infection. *New England Journal of Medicine* **322**, 941–949.
- Wei, G. C. G. and Tanner, M. A. (1990). Posterior computations for censored regression data. *Journal of the American Statistical Association* **85**, 829–839.
- Wei, G. C. G. and Tanner, M. A. (1991). Applications of multiple imputation to the analysis of censored regression data. *Biometrics* **47**, 1297–1309.
- Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.
4. Missingness of the longitudinal auxiliary variable, in the sense that  $\mathbf{y}$  is observed instead of  $\mathbf{y}^f$ , is ignorable (Rubin, 1976).
- Under these assumptions, using arguments analogous to those used in the appendix of Faucett et al. (1998), it can be shown that our imputation procedure is equivalent to drawing  $\mathbf{T}^0$  from the correct posterior predictive distribution with density  $p(\mathbf{T}^0 | \mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{D})$ . Implicit in this result is an assumption that, given  $\mathbf{x}$  and  $\mathbf{z}$ , the distribution of  $T^0$  does not depend on  $\mathbf{y}^f$ . A similar assumption was made in Section 2.3.2.
- The parameters of primary interest, say  $\gamma^0$ , are the parameters of  $[T^0 | \mathbf{x}^0]$ . Under assumptions 1–4, the posterior distribution of  $\gamma^0$  can be shown to be equal to
- $$\int p(\gamma^0 | \mathbf{T}^0, \mathbf{X}, \mathbf{Y}) p(\mathbf{T}^0 | \mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{D}) d\mathbf{T}^0. \quad (1)$$
- Moreover, suppose that a fifth assumption holds, i.e.,
5. The parameters of  $[\mathbf{y}, \mathbf{x}^1 | T^0, \mathbf{x}^0]$  are distinct from  $\gamma^0$ , where  $\mathbf{x}^1$  contains the variables in  $\mathbf{x}$  other than  $\mathbf{x}^0$ .
- Then it can be shown that
- $$[\gamma^0 | \mathbf{T}^0, \mathbf{X}, \mathbf{Y}] = [\gamma^0 | \mathbf{T}^0, \mathbf{X}^0], \quad (2)$$
- and (1) becomes
- $$\int p(\gamma^0 | \mathbf{T}^0, \mathbf{X}^0) p(\mathbf{T}^0 | \mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{D}) d\mathbf{T}^0. \quad (3)$$
- Our MI inference procedure uses simulation to approximate (3), with  $p(\gamma^0 | \mathbf{T}^0, \mathbf{X}^0)$  approximated by the primary analysis. Thus, to the extent that the primary analysis provides an adequate approximation (see Section 2.5) and assumptions 1–5 hold, our procedure is valid under our joint model for the observed data.

Received March 2000. Revised August 2001.

Accepted August 2001.

## APPENDIX

### Conditions for Validity of the Procedures

We describe conditions for the validity of our procedures when we do not impose time limits on the imputed values of  $T^0$ . In Remark C, we discuss the effect of imposing such time limits.

Let  $\mathbf{y}^f$  denote the full history of the longitudinal auxiliary variable up to time  $T$  and let  $\mathbf{y}^m$  denote the values in  $\mathbf{y}^f$  that are not included in the observed  $\mathbf{y}$  due to only periodic follow-up. The assumptions about censoring and missing data that underlie our imputation procedure are

1.  $T^0$  and  $C$  are independent, given  $\mathbf{x}$ ,  $\mathbf{y}^f$ , and  $\mathbf{z}$ .
2.  $C$  is independent of  $\mathbf{y}^m$  and  $\mathbf{z}$ , given  $\mathbf{x}$  and  $\mathbf{y}$ .
3. The parameters of  $[T^0, \mathbf{y}^f, \mathbf{z} | \mathbf{x}]$  are distinct from the parameters of  $[C | \mathbf{x}, \mathbf{y}^f, \mathbf{z}]$ .

*Remark A.* Standard assumptions about censoring for the PO analysis are that  $T^0$  and  $C^0$  are independent given  $\mathbf{x}^0$  and that the parameters of  $[T^0 | \mathbf{x}^0]$  are distinct from the parameters of  $[C | \mathbf{x}^0]$  (Kalbfleisch and Prentice, 1980, p. 40). It can be shown that the former (independence) assumption is equivalent to the following:

- 1'.  $T^0$  and  $C$  are independent given  $\mathbf{x}^0$  and  $\mathbf{z}$ .
- 2'. Either  $T^0$  and  $\mathbf{z}$  are independent given  $\mathbf{x}^0$  or  $C$  and  $\mathbf{z}$  are independent given  $\mathbf{x}^0$ .

Assumption 1' is usually stronger than the corresponding MI Assumption 1 because the former assumes independence with less conditioning than the latter. Moreover, as long as the latent auxiliary variables  $\mathbf{z}$  contain information about  $T^0$  in addition to that contained in  $\mathbf{x}^0$ , assumption 2' becomes an assumption that  $C$  and  $\mathbf{z}$  are independent given  $\mathbf{x}^0$ , which similarly is often stronger than Assumption 2 with regard to  $C$  and  $\mathbf{z}$ .

*Remark B.* Assumption 5 underlying our MI procedure and the resulting equation (2) express the notion that  $\mathbf{y}$  and  $\mathbf{x}^1$  do not contain any information about  $\gamma^0$  in addition to that contained in  $T^0$  and  $\mathbf{x}^0$ . This assumption is difficult to check for our joint model. We suspect that violation of the assumption

is analogous to the second type of incompatibility discussed in Section 2.5, in which the imputation model incorporates extra information that is not used in the procedure for analyzing the augmented data.

*Remark C.* Because we censor our imputations at  $t_{\max}$  days (see Section 2.4.2), our method's adjustment for dependent censoring is only a partial adjustment in some cases. For example, consider the situation in which subjects with higher CD4 are censored more frequently and also tend to have longer failure times. (This type of situation was dis-

cussed briefly in Section 3.1.) Our method adjusts for such dependent censoring by imputing longer failure times for censored subjects with higher CD4 counts. But our censoring at  $t_{\max}$  days can counteract this adjustment for subjects with imputed failure times that are longer than  $t_{\max}$  days. As the discussions in Section 2.4.2 and this remark imply, our imputation procedure in a sense balances not doing a complete job of adjusting for dependent censoring against extrapolating beyond the range of the data.