Research Article

Statistics in Medicine

Received 14 November 2015, Accepted 27 June 2016 Published online 21 July 2016 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7048

Multiple imputation of missing covariates for the Cox proportional hazards cure model

Lauren J. Beesley,[a*†] Jonathan W. Bartlett,[b] Gregory T. Wolf[c] and Jeremy M. G. Taylor[a]

We explore several approaches for imputing partially observed covariates when the outcome of interest is a censored event time and when there is an underlying subset of the population that will never experience the event of interest. We call these subjects 'cured', and we consider the case where the data are modeled using a Cox proportional hazards (CPH) mixture cure model. We study covariate imputation approaches using fully conditional specification. We derive the exact conditional distribution and suggest a sampling scheme for imputing partially observed covariates in the CPH cure model setting. We also propose several approximations to the exact distribution that are simpler and more convenient to use for imputation. A simulation study demonstrates that the proposed imputation approaches outperform existing imputation approaches for survival data without a cure fraction in terms of bias in estimating CPH cure model parameters. We apply our multiple imputation techniques to a study of patients with head and neck cancer. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: multiple imputation; cure models; fully conditional specification

1. Introduction

In survival analysis, a common assumption is that all subjects will eventually experience the event of interest given long enough follow-up time. However, there are many settings in which this assumption does not hold. For example, suppose we are interested in studying cancer recurrence in patients treated for head and neck cancer. If the treatment completely eradicated the cancer in some individuals, then there will be a subset of the population that will never experience a recurrence. We call these subjects 'cured' or 'non-susceptible'.

One commonly used modeling approach for survival data with a cured fraction is a mixture model with two components. The first component is a model for the probability that a subject is not cured, which is usually modeled using logistic regression. The second component is a model for the failure time in the susceptible (non-cured) population. Parametric, semiparametric, and nonparametric formulations of the failure time model exist in the literature [1–7]. We consider a formulation of the mixture cure model where failure time in the susceptible population is modeled using a Cox proportional hazards (CPH) regression model [4, 6, 8]. It is important to note that subjects with observed events are known to be non-cured, but cure status is not known for censored subjects. Cure models are appealing because they enable enhanced interpretation and inference from data with a cure structure as cure models allow us to model both the probability that a subject is cured and the hazard of an event in the non-cured group separately.

A challenge that arises in the application of these cure models is that often one or more covariates are only partially observed. One simple approach is to ignore the missing data and analyze only the patients with complete covariate data. 'Complete case' (CC) analysis is an undesirable approach because it does not use data from patients with missing covariate values and is therefore inefficient. Also, CC analysis

[a]*Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*
[b]*Statistical Innovation Group, AstraZeneca, Cambridge, U.K.*
[c]*Department of Otolaryngology, University of Michigan, Ann Arbor, MI, U.S.A.*
[*]*Correspondence to: Lauren J. Beesley, Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*
[†]*E-mail: lbeesley@umich.edu*

may be biased if the covariate missingness mechanism depends on the outcome. Other approaches in the literature for handing missing covariates in the cure setting often involve modeling the joint distribution of the missing covariates using general location models [7, 9] or by specifying a series of conditional distributions [10]. Both approaches require us to explicitly specify the joint distribution of the covariates, which may not be easily carried out, and they are not easily implemented using standard software.

In this paper, we explore multiple imputation as another approach for handling missing data in the cure model setting. When performing multiple imputation, it is important to include outcome information in the model for imputing partially observed covariates [11]. In the cure setting, however, many aspects of the outcome (cure status and event times in the non-cured subjects) are not fully observed because of censoring. We are interested in comparing different methods for incorporating the observed outcome information to impute partially observed covariates when the primary outcome has a CPH cure structure. We will study covariate imputation approaches using fully conditional specification (FCS).

Fully conditional specification is a multiple imputation approach in which we specify a conditional distribution for each partially observed covariate [12, 13]. We then use these conditional distributions to impute covariates as part of an iterative algorithm that cycles through the conditional distributions for all the partially observed covariates. This often involves specifying a regression model for each partially observed covariate and then using the regression models to impute the missing values. An attractive feature of FCS is that it does not require us to explicitly specify the joint distribution of the covariates. Suppose X is a set of covariates and Y is an outcome variable. Also, suppose our ultimate goal is to fit a standard regression model for $Y|X$ (e.g., linear, logistic). Let $X^{(p)}$ denote the p^{th} covariate in X and $X^{(-p)}$ denote all covariates in X except $X^{(p)}$. We would like to use the distribution of $X^{(p)}|X^{(-p)}, Y$ to impute each partially observed $X^{(p)}$. If we have the distributions for $Y|X$ and $X^{(p)}|X^{(-p)}$, then we can derive the distribution for $X^{(p)}|X^{(-p)}, Y$ directly. When $X^{(p)}|X^{(-p)}$ and $Y|X$ are normally distributed with predictors incorporated in the mean structure, then the distribution of $X^{(p)}|X^{(-p)}, Y$ will also be normal and will correspond to a linear regression that can be readily used to impute $X^{(p)}$. When the true distribution of $X^{(p)}|X^{(-p)}, Y$ is unknown or difficult to sample from, we may attempt to approximate the distribution using a simpler and more computationally convenient standard regression model. For example, for normal $X^{(p)}$, we may specify the distribution of $X^{(p)}|X^{(-p)}, Y$ using some function of $X^{(-p)}$ and Y as predictors in a linear regression model.

In survival analysis, the primary outcome usually consists of the pair (Y, δ). If T is the underlying event time and C is the censoring time, then $Y = \min(T, C)$ and $\delta = I(T \leqslant C)$. The ultimate goal is usually to fit a model for $T|X$. Although T is the outcome of interest, it is not directly observed because of censoring. We can still derive the exact distribution of $X^{(p)}|X^{(-p)}, Y, \delta$ to impute each partially observed $X^{(p)}$. However, because of the complicated structure of survival data, the exact distribution of $X^{(p)}|X^{(-p)}, Y, \delta$ will often be inconvenient or computationally intensive to sample from [14].

One possible alternative is to obtain a more convenient approximation to the exact conditional distribution of $X^{(p)}|X^{(-p)}, Y, \delta$ for each partially observed covariate $X^{(p)}$. White and Royston derived an approximate conditional distribution for proportional hazards survival data that reduced to a regression model of $X^{(p)}$ with predictors $X^{(-p)}$, δ, and $\hat{H}_0(Y)$, where $\hat{H}_0(Y)$ is the estimated cumulative baseline hazard function [15]. One adaptation of this would be to using log(Y) in place of $\hat{H}_0(Y)$ [16]. Another adaptation would be to use a regression model for $X^{(p)}$ with predictors $X^{(-p)}$, $\delta f_1(Y)$, and $(1-\delta)f_2(Y)$, where $f_1(Y)$ and $f_2(Y)$ are functions of Y specified using splines or step functions.

Additionally, because $Y = \min(T, C)$ is a mixture of a censoring time and the event time of interest, it may not be appealing to include Y in the imputation regression models, and we may instead wish to incorporate T directly. We can treat T as another partially observed variable and impute the value of T from the distribution of $T|T > C, X$ for censored subjects. Assuming C is uninformative for $X^{(p)}$, we can then try to impute each partially observed $X^{(p)}$ by specifying the exact conditional distribution $X^{(p)}|X^{(-p)}, T$ or by approximating the exact distribution with a regression model using T.

When the ultimate goal is to fit a mixture cure model, the form for the distribution of $T|X$ is more complicated. The most convenient estimation method introduces a partially observed variable, G, which indicates cure status. Either an imputed value or the expectation of G is used in the mixture cure model estimation algorithm [6]. When we have partially observed covariates, we can impute each partially observed $X^{(p)}$ from the corresponding distribution of $X^{(p)}|X^{(-p)}, Y, \delta, G$. Using assumptions for the distribution of $X^{(p)}|X^{(-p)}$, we can derive the exact conditional distribution from which to impute. We can also impute using approximations to the exact conditional distribution that are more computationally convenient. Alternatively, we can impute the event time T for censored individuals and then impute each partially observed $X^{(p)}$ using the approximated conditional distribution of $X^{(p)}|X^{(-p)}, T, G$.

In this paper, we derive the exact conditional distribution and suggest a sampling scheme for imputing partially observed covariates in the CPH mixture cure model setting. Additionally, we propose several approximations to the exact distribution that are more convenient to use for imputation. We compare the performance of our proposed imputation approaches with methods for survival data without a cure fraction.

In Section 2, we present details about the CPH cure model. In Section 3, we present possible approaches for imputing partially observed covariates in the cure setting. In Section 4, we report results from a set of simulations and compare the performance of the imputation algorithms. In Section 5, we apply two imputation approaches to a study of cancer recurrence in head and neck cancer patients, and in Section 6, we present a discussion.

2. Cox proportional hazards cure model

We consider the setting where the primary outcome is a censored event time, and there is an underlying subset of the study population that will never experience the event of interest. We call individuals that will never experience the event 'cured'. The CPH cure model is a mixture model with two components: (1) a model for the probability that an individual is not cured; and (2) a CPH model for the hazard of an event for non-cured subjects [4].

Let $Y_i = \min(T_i, C_i)$ be the observed event/censoring time for individual i where T_i is the underlying event time (defined as infinity if a subject is cured) and C_i is the censoring time. Let $\delta_i = I(T_i \leqslant C_i)$. We define the cure status of individual i, G_i, as 1 when the individual is not cured and 0 when the individual is cured. G_i is 1 when $\delta_i = 1$ and is unknown when $\delta_i = 0$. We assume censoring is independent of G and T given covariates. We model the data as follows:

$$\text{Logistic Model of Cure Status: } \text{logit}(P(G_i = 1|X_i)) = \alpha_0 + \alpha^T X_i \; i = 1, ..., n$$
$$\text{CPH Model of Failure Time: } h(t|X_i, G_i = 1) = h_0(t)e^{\beta^T X_i} \; i = 1, ..., n$$

where $h_0(t)$ is the baseline hazard of having an event in the non-cured group. For simplicity, we assume that we have the same set of covariates in both parts of the mixture model. Estimation of model parameters can be carried out using an expectation–maximization (EM) algorithm [5, 6].

We consider the complete data partial log-likelihood corresponding to the CPH cure model assuming that G_i is observed. The EM algorithm iterates between two steps. In the E-step for a given iteration, we replace G_i in the complete data partial log-likelihood with

$$w_i = E(G_i|\delta_i, Y_i, X_i) = \delta_i + (1 - \delta_i)\frac{p_i S(Y_i|X_i, G_i = 1)}{1 - p_i + p_i S(Y_i|X_i, G_i = 1)} \tag{1}$$

Here, $p_i = P(G_i = 1|X_i) = \text{expit}(\alpha_0 + \alpha^T X_i)$ and $S(Y_i|X_i, G_i = 1) = e^{-H_0(Y_i)e^{\beta^T X_i}}$ using the estimates of α_0, α, and β from the previous iteration and an estimate of $H_0(t)$ obtained using a Breslow estimator weighted by w_i [17]. To improve the stability of the EM algorithm (model parameters are nearly unidentifiable), we define censored individuals with very late censoring times as cured with $w_i = 0$ [6]. The M-step involves taking the complete data partial log-likelihood with w_i substituted for G_i and maximizing it with respect to α_0, α, and β. The EM algorithm allows us to handle the fact that cure status is only partially observed. Variances of model parameter estimates can be estimated via bootstrap.

3. Multiple imputation of missing covariates

In this section, we discuss imputation by FCS in more detail. Then, we derive the exact conditional distribution to impute partially observed covariates in the cure setting. We also present several approximations to the exact distribution that are more convenient to use for imputation. We include several covariate imputation models for survival data without a cured fraction.

3.1. Fully conditional specification

Fully conditional specification or 'chained equations' is a multiple imputation approach in which we specify the conditional distribution for each partially observed variable and then use these distributions to impute variables one-by-one as part of an iterative procedure [12, 13]. Suppose we are interested in fitting a model to outcome O with partially observed covariates $W = (X^{(1)}, \ldots, X^{(d)})$ and fully observed covariates $Z = (X^{(d+1)}, \ldots, X^{(s)})$. Let $X = (W, Z)$. Recall that $X^{(p)}$ denotes the p^{th} covariate in X and $X^{(-p)}$

denotes all covariates in X except $X^{(p)}$. For each partially observed $X^{(p)}$, we specify the conditional distribution $f(X^{(p)}|X^{(-p)}, O; \phi^p)$ where ϕ^p is a set of parameters. Let $f(\phi^p|X, O)$ denote the posterior distribution of ϕ^p and let $X^{(p,miss)}$ and $X^{(p,obs)}$ denote the missing and observed portions of $X^{(p)}$. To impute missing values for $X^{(1)} \dots X^{(d)}$, we perform the following iterative chained equations algorithm. At iteration k, we obtain updated imputed values by drawing

$$\begin{aligned}
\phi^1_{(k)} &\sim f\left(\phi^1|X^{(1,obs)}_{(k-1)}, \dots, X^{(d)}_{(k-1)}, Z, O\right)\\
X^{(1,miss)}_{(k)} &\sim f\left(X^{(1)}|X^{(2)}_{(k-1)}, \dots, X^{(d)}_{(k-1)}, Z, O; \phi^1_{(k)}\right)\\
\phi^2_{(k)} &\sim f\left(\phi^2|X^{(1)}_{(k)}, X^{(2,obs)}_{(k-1)}, \dots, X^{(d)}_{(k-1)}, Z, O\right)\\
X^{(2,miss)}_{(k)} &\sim f\left(X^{(2)}|X^{(1)}_{(k)}, X^{(3)}_{(k-1)}, \dots, X^{(d)}_{(k-1)}, Z, O; \phi^2_{(k)}\right)\\
&\dots\\
\phi^d_{(k)} &\sim f\left(\phi^d|X^{(1)}_{(k)}, \dots, X^{(d-1)}_{(k)}, X^{(d,obs)}_{(k-1)}, Z, O\right)\\
X^{(d,miss)}_{(k)} &\sim f\left(X^{(d)}|X^{(1)}_{(k)}, \dots, X^{(d-1)}_{(k)}, Z, O; \phi^d_{(k)}\right)
\end{aligned}$$

We iterate until convergence. When we have missingness in only one variable, no iteration is required, and the algorithm reduces to standard parametric multiple imputation.

In our cure setting, we want to use the conditional distribution $f(X^{(p)}|X^{(-p)}, Y, \delta, G; \phi^p)$ to impute each partially observed covariate $X^{(p)}$. In practice, however, $f(X^{(p)}|X^{(-p)}, Y, \delta, G; \phi^p)$ may be difficult to use for imputation, and we may use an approximation, $\tilde{f}(X^{(p)}|X^{(-p)}, Y, \delta, G; \tilde{\phi}^p)$. We approximate the posterior distribution of $\tilde{\phi}^p$ (or ϕ^p) by a multivariate normal distribution. If the distribution used for imputation explicitly depends on G, we treat G as another partially observed variable and impute G as part of the chained equations algorithm. If we also impute the true event time T for censored subjects, we could impute partially observed $X^{(p)}$ using $f(X^{(p)}|X^{(-p)}, T, G; \phi^p)$ or a corresponding approximation. We assume that the covariates are missing at random (MAR).

For many of the imputation approaches we consider, drawing $\tilde{\phi}^p$, and missing $X^{(p)}$ values will reduce to first fitting a regression model for $X^{(p)}$ using some function of $X^{(-p)}$, G, Y, δ, and maybe T as predictors. As in standard FCS, we fit this regression model only for subjects with observed $X^{(p)}$. We then draw the parameter $\tilde{\phi}^p$ from a multivariate normal with mean and variance obtained using the regression model fit and use the drawn $\tilde{\phi}^p$ and the conditional distribution implied by the regression model to draw each missing value of $X^{(p)}$. We will call this regression model the imputation model for $X^{(p)}$. Alternatively, we can obtain a draw of $\tilde{\phi}^p$ by fitting the imputation model to a bootstrap sample of the data [18]. Multiple imputation using standard regression models can be implemented using the package MICE in R [19]. For imputing covariates assumed to be normally distributed, we use predictive mean matching as implemented in MICE.

The chained equations (FCS) algorithm will result in a single imputed dataset. We repeat the algorithm to create several imputed datasets. Suppose our goal is to make inference from a particular model fit (in our case, the CPH cure model). We fit this model to each imputed dataset, and then we use Rubin's Rules to produce a final estimate of the parameters and their variances from which we can make the desired inference [20].

3.2. Imputation using the exact conditional distribution

We can use the complete data likelihood from the CPH cure model and an assumption about the distribution of $X^{(p)}|X^{(-p)}$ to derive the kernel of the conditional distribution of $X^{(p)}|X^{(-p)}, \delta, G$, and Y for each partially observed $X^{(p)}$.

Below, we derive the exact imputation distribution assuming $X^{(p)}_i \sim N(\theta_0 + \theta^T X^{(-p)}_i, \sigma^2)$. We can generalize our approach to impute covariates with non-normal distributions. We include a derivation for Bernoulli random variables in the appendix. We assume that censoring does not depend on $X^{(p)}$ but may depend on other covariates. Therefore, we do not need to specify a model for the censoring mechanism to derive the conditional distribution of $X^{(p)}$. Let $f(X^{(-p)}_i; \gamma)$ be the joint distribution of $X^{(-p)}_i$. In practice, we will not need to explicitly specify this distribution. Let $f(X^{(p)}_i|X^{(-p)}_i; \theta_0, \theta, \sigma^2)$ be the distribution of $X^{(p)}$

given all the other covariates. We consider the complete data likelihood (assuming cure status is known) for the CPH cure model:

$$L(\alpha,\alpha_0,\beta,\theta,\theta_0,\gamma,\sigma^2)=\prod_{i=1}^{n}\left\{h(Y_i|G_i=1,X_i;\beta)^{\delta_i}S(Y_i|G_i=1,X_i;\beta)P(G_i=1|X_i;\alpha,\alpha_0)\right\}^{G_i}$$
$$\times\left\{P(G_i=0|X_i;\alpha,\alpha_0)\right\}^{1-G_i}f\left(X_i^{(p)}|X_i^{(-p)};\theta_0,\theta,\sigma^2\right)f\left(X_i^{(-p)};\gamma\right)$$

$$\propto\prod_{i=1}^{n}\left\{\left(h_0(Y_i)e^{\beta^TX_i}\right)^{\delta_i}e^{-H_0(Y_i)e^{\beta^TX_i}}\frac{e^{\alpha^TX_i+\alpha_0}}{1+e^{\alpha^TX_i+\alpha_0}}\right\}^{G_i}\left\{\frac{1}{1+e^{\alpha^TX_i+\alpha_0}}\right\}^{1-G_i}e^{\frac{-\left(X_i^{(p)}-\theta_0-\theta^TX_i^{(-p)}\right)^2}{2\sigma^2}}f\left(X_i^{(-p)};\gamma\right)$$

From the aforementioned likelihood, we see that

$$f\left(X_i^{(p)}|G_i,\delta_i,Y_i,X_i^{(-p)}\right)\propto\left\{e^{\delta_i\beta^TX_i}e^{-H_0(Y_i)e^{\beta^TX_i}}\frac{e^{\alpha^TX_i+\alpha_0}}{1+e^{\alpha^TX_i+\alpha_0}}\right\}^{G_i}\left\{\frac{1}{1+e^{\alpha^TX_i+\alpha_0}}\right\}^{1-G_i}e^{\frac{-\left(X_i^{(p)}-\theta_0-\theta^TX_i^{(-p)}\right)^2}{2\sigma^2}}\quad(2)$$

We can use this kernel to draw from $f(X_i^{(p)}|G_i,\delta_i,Y_i,X_i^{(-p)})$ within the chained equations imputation procedure. We note that this kernel depends on both G_i and $H_0(t)$, and it is parameterized by $\alpha,\alpha_0,\beta,\sigma^2,\theta$, and θ_0. When $X_i^{(p)}$ is assumed to be normal, we can draw from (2) using an accept–reject algorithm as described hereafter. When $X_i^{(p)}$ is binary and modeled as in the appendix, we do not require an accept–reject algorithm to draw from $f(X_i^{(p)}|G_i,\delta_i,Y_i,X_i^{(-p)})$.

In order to impute partially observed covariates using their exact conditional distributions, we treat G as another partially observed variable and impute G within the chained equations algorithm. We also append a step at the start of each chained equations iteration in which we estimate $H_0(t)$. We can impute by iterating the following steps:

Step 1: Estimating $H_0(t)$

We can estimate $H_0(t)$ several different ways. Firstly, we can estimate $H_0(t)$ using a weighted Breslow estimator [17]. Suppose we have event times $t_1,...,t_J$ and let R_j be the risk set at time t_j. Using the imputed X from the most recent iteration, we estimate $H_0(t)$ at the k^{th} iteration of the imputation algorithm as the step function

$$\hat{H}_0^{(k)}(t)=\sum_{t_j\leqslant t}^{J}\frac{\#\text{ events at time }t_j}{\sum_{i\in R_j}e^{\left[\beta^{(k-1)}\right]^TX_iw_i^{(k)}}}$$

where $w_i^{(k)}$ is the conditional probability that a person is not cured at iteration k as expressed in Equation (1) and $\beta^{(k-1)}$ is a draw of β from the previous iteration [6]. We use this approach to estimate $H_0(t)$ in our simulations.

We can also obtain a parametric estimate of $H_0(t)$ by fitting a CPH cure model with a parametric baseline hazard such as Weibull. If the baseline hazard of an event in the non-cured subjects is truly Weibull, then fitting a Weibull cure model rather than a semi-parametric CPH cure model may produce extra efficiency in estimating β. However, if the baseline hazard in the non-cured group is not believed to be Weibull, using this approach is not advised. Alternatively, $H_0(t)$ can be estimated using only the subset of the data such that $G_i=1$ (non-cured) as imputed at iteration $k-1$. This can be estimated by fitting a Cox model and using a traditional Breslow estimator applied to the $G_i=1$ subset of the data or by assuming a parametric form for the event hazard in the $G_i=1$ group.

Step 2: Imputing cure status

To produce proper imputations using the FCS algorithm, we first draw the parameters. We can obtain draws of α_0,α, and β at a given iteration by (1) fitting a logistic model to the most recent imputed data with G as the outcome and X as covariates, (2) fitting a CPH regression model to the subset of subjects such that $G_i=1$, and then (3) drawing (α_0,α,β) from a multivariate normal distribution using the estimated parameters and their corresponding covariance matrices from the logistic and CPH model fits. This approach is much faster than fitting a cure model to the data to estimate the parameters and then using bootstrap to estimate the covariance matrix. Alternatively, we can draw (α_0,α,β) by fitting the models in (1) and (2) to a bootstrap sample [18].

Using the complete data likelihood for the CPH cure model, we can show that $\text{logit}(P(G_i = 1|X_i, \delta_i = 0, Y_i)) = -\hat{H}_0(Y_i)e^{\beta^T X_i} + \alpha^T X_i + \alpha_0$. We can draw imputed values of G_i using this probability relation. We note that if $\delta_i = 1$, then G_i is known to be 1, so we will not need to impute. Also, we define censored individuals with late censoring times (after some cut-point c) as cured. Therefore, G is treated as missing only if $\delta = 0$ and $Y \leqslant c$, so we can view missingness in G as MAR conditional on δ and Y.

Step 3: Imputing the missing covariates

We specify the distribution $f(X_i^{(p)}|G_i, \delta_i, Y_i, X_i^{(-p)}; \phi^p)$ for each covariate $X^{(p)}$ with missing values. As described in Section 3.1, we (1) draw ϕ^p and (2) impute missing values of $X^{(p)}$ for each $X^{(p)}$ in $X^{(1)}, \ldots, X^{(d)}$. If only one covariate has missingness, we perform (1) and (2) a single time for that covariate. If we have missingness in many covariates, we perform (1) and (2) sequentially for each covariate with missingness using the most recent imputations of the other variables. We describe how to perform (1) and (2) to impute normal and binary covariates using their exact conditional distributions.

Normal $X^{(p)}$: We can draw $(\theta_0, \theta, \sigma^2)$ under the Bayesian linear regression model with $X^{(p)}$ as the outcome and with $X^{(-p)}$ as the predictors using the most recent imputed values. This model is described by Rubin (1987) [20] and used in MICE [19]. Unlike standard FCS, we fit this model using all subjects and the complete imputed $X^{(p)}$ from the most recent iteration as the outcome [14]. If desired, we may also draw new values of α and β as described in Step 2 and using the newly-imputed G. We then want to impute each missing value $X_i^{(p)}$ by taking draws from the full conditional distribution knowing only the kernel in (2). Many methods exist to draw from a distribution using only the kernel. To obtain an imputed value for $X_i^{(p)}$ at a given iteration, we perform a Metropolis–Hastings draw from (2) using a normal random walk proposal distribution centered at the imputed value from the previous iteration [21, 22]. The variance of this proposal distribution is a tuning parameter that must be determined to ensure good mixing properties and a reasonable acceptance rate [23]. Because of this accept–reject sampling, we may need to perform many iterations of the chained equations fitting algorithm to reach convergence.

Binary $X^{(p)}$: Using notation from the appendix, we draw (θ_0, θ) using a logistic regression fit with $X^{(p)}$ as the outcome and $X^{(-p)}$ as covariates. We then impute missing values $X_i^{(p)}$ using the probability relation in Equation (4). This reduces to drawing $X_i^{(p)}$ from a Bernoulli(π_i) distribution with parameter $\pi_i = P(X_i^{(p)} = 1|G_i, \delta_i, Y_i, X_i^{(-p)})$ from (A.1).

This 'Exact Cure' approach imputes each partially observed $X^{(p)}$ using its conditional distribution implied by the CPH cure model and the model for $X^{(p)}|X^{(-p)}$. However, when $X^{(p)}$ is normal, sampling from this specification of $f(X_i^{(p)}|G_i, \delta_i, Y_i, X_i^{(-p)})$ requires us to use an accept–reject algorithm to impute each missing $X_i^{(p)}$ at each iteration of the chained equations imputation procedure, and this can quickly result in a large computational burden. This burden is amplified when we have missingness in multiple covariates. To impute multiple partially observed covariates, we must specify the model for $X^{(p)}|X^{(-p)}$ for each partially observed $X^{(p)}$, which increases the number of parameters that must be drawn. Additionally, we must derive the form of $f(X_i^{(p)}|G_i, \delta_i, Y_i, X_i^{(-p)})$ separately for different forms of the model for each $X^{(p)}|X^{(-p)}$ (e.g., Gamma, Poisson, etc). Because of this, we do not apply the Exact Cure approach to the head and neck cancer example later on, which has missingness in many variables.

3.3. Approximations to the exact distribution using regression models

In the previous section, we derived exact conditional distributions to use for imputation of normal and binary covariates and sampling from these distribution can often become computationally intensive. We will consider approximations to the exact conditional distributions that do not require accept–reject sampling and can more easily be implemented with existing software. We are interested in approximations that correspond to standard regression models. We can then perform a FCS draw from the approximate conditional distribution by fitting a standard regression model as described earlier.

We start by describing two simple covariate imputation approaches for survival data without a cure fraction. We then describe an approach in the literature for imputing survival data without a cure fraction that is motivated directly by the standard CPH model. Then, we propose an approximate distribution that incorporates the cure structure of the data and is motivated by the CPH cure model formulation. Finally, we consider a modification to these approaches in which event time T is imputed for censored subjects.

3.3.1. logY imputation for survival data without a cure fraction. One approach in the literature for imputing covariates for survival data without a cure fraction is to use $X^{(-p)}$, δ, and $\log(Y)$ as predictors in the imputation model for $X^{(p)}$ used in the chained equations algorithm [16]. Unlike the Exact Cure approach, this approach does not require us to impute cure status or estimate $H_0(t)$, so we do not require iteration of the chained equations algorithm when we have missingness in only one covariate. We can impute using MICE in R by specifying regression models with predictors $X^{(-p)}$, δ, and $\log(Y)$ for imputing each partially observed $X^{(p)}$ [19].

3.3.2. Outcome binning imputation for survival data without a cure fraction. One adaptation of existing approaches for imputing covariates in the non-cure setting would be to use a regression model for imputing each partially observed $X^{(p)}$ with predictors $X^{(-p)}$, $\delta f_1(Y)$ and $(1-\delta)f_2(Y)$ where $f_1(Y)$ and $f_2(Y)$ are some functions of Y. We propose using f_1 and f_2 in the form of step functions with step height determined by the data. This allows for a very flexible association between the outcome and the partially observed covariate. Additionally, this approach does not require us to impute cure status or estimate $H_0(t)$ explicitly.

We call this approach 'Outcome Binning' because it involves binning individuals based on the composite outcome, (Y, δ). We first separate subjects into a $\delta = 1$ and $\delta = 0$ group. We then define bins of Y within each δ group using summary statistic-based cutoffs or by other methods. For convenience, we define the bins using quartiles of Y within each of the $\delta_i = 1$ and $\delta_i = 0$ groups. We define a set of dummy indicator variables, $M_1, \ldots, M_m$, which identify the bin membership of each individual ($M_k = 1$ if the subject is in bin k). We then impute each partially observed covariate within the chained equations procedure using a regression model for each $X^{(p)}$ with $X^{(-p)}$ and binary indicators $M_2, \ldots, M_m$ as predictors. After determining $M_1, \ldots, M_m$, we can perform the chained equations imputation using MICE in R [19]. With missingness in only one covariate, we can perform a single iteration of the chained equations algorithm.

3.3.3. White and Royston imputation for the Cox proportional hazards model without a cure fraction. Based on algebraic derivation involving Taylor approximations, White and Royston suggest using $X^{(-p)}$, δ, and $H_0(Y)$ as predictors in the imputation model for each partially observed $X^{(p)}$ in the standard CPH model setting without a cure fraction [15]. This is quite similar to the approximation in Section 3.3.1 but replacing $\log(Y)$ with $H_0(Y)$. This requires us to obtain an estimate of $H_0(t)$ but does not require us to impute cure status.

We note that $H_0(t)$ is the cumulative baseline hazard of an event in the entire study population. This is not the same as the cumulative baseline hazard in the non-cured population, as the cured subjects cannot experience the event of interest. When applied to survival data with a cure fraction, $H_0(t)$ is the cumulative baseline hazard of an event in the (assumed to be misspecified) survival model without a cure fraction based on the entire study population.

White and Royston ultimately recommend using the Nelson–Aalen estimator of $H(t)$ to estimate $H_0(t)$ before imputation. However, they also investigated an approach in which they add a step to the imputation algorithm and re-estimate $H_0(t)$ at each iteration. We estimate $H_0(t)$ after each iteration of the chained equations algorithm by fitting a Cox model to all subjects using the most recent imputed data, drawing the Cox model parameter using a multivariate normal distribution with mean and covariance matrix from the Cox model fit, and then using a Breslow estimator. We can also draw parameter values by fitting the models to a bootstrap sample of the data [18]. Alternatively, we can fit a Weibull regression model to all subjects and estimate the cumulative baseline hazard in the total population as a parametric function.

As we estimate $H_0(t)$ at the end of each iteration, we iterate the chained equations algorithm even when we only have missingness in a single covariate. We can impute using MICE in R by iterating the following steps: (1) estimate $H_0(t)$ and (2) impute each partially observed covariate $X^{(p)}$ sequentially using an appropriate elementary imputation method in MICE (e.g., mice.impute.logreg() for binary covariates) with predictors $X^{(-p)}$, δ, and $\hat{H}_0(Y)$ [19].

3.3.4. Approximated imputation for the Cox proportional hazards cure model. We use a similar approach to White and Royston to derive approximate imputation models for normal and binary covariates in the CPH cure model setting [15]. Although not shown here, we can derive approximate imputation models for covariates with other distributions in a similar fashion. Suppose we have the same set of covariates in both parts of the mixture cure model and that the set contains s covariates. Therefore, α and β both have dimension s. Again, we suppose that a partially observed $X^{(p)} \sim N(\theta^T X^{(-p)} + \theta_0, \sigma^2)$. Taking the logarithm of kernel (2), we have that

$$\log\left(f\left(X_i^{(p)}|G_i,\delta_i,Y_i,X_i^{(-p)}\right)\right) = \frac{-1}{2\sigma^2}\left(X_i^{(p)} - \theta^T X_i^{(-p)} - \theta_0\right)^2 + G_i\alpha^T X_i - \log\left(1+e^{\alpha^T X_i+\alpha_0}\right) + G_i\delta_i\beta^T X_i + G_i\delta_i\log\left(h_0(Y_i)\right) - G_i H_0(Y_i)e^{\beta^T X_i} + \text{constant}$$

We treat terms that do not depend on $X_i^{(p)}$ as constant. We note that log(1+z) ≈ log(1+c) + (z-c)/(1+c) if z is near c and $e^{aX+bY} \approx e^{a\bar{X}+b\bar{Y}}\left[1+a(X-\bar{X})+b(Y-\bar{Y})\right]$ if Var(aX + bY) is small. Assuming Var($\alpha^T X_i$) and Var($\beta^T X_i$) are small, we can approximate the above by:

$$\begin{aligned}\log\left(f\left(X_i^{(p)}|G_i,\delta_i,Y_i,X_i^{(-p)}\right)\right) &\approx \frac{-1}{2\sigma^2}\left(X_i^{(p)} - \theta^T X_i^{(-p)} - \theta_0\right)^2 + G_i\alpha^T X_i \\ &\quad - \frac{e^{\alpha^T\bar{X}+\alpha_0}}{1+e^{\alpha^T\bar{X}+\alpha_0}}\left[1+\alpha_p\left(X_i^{(p)}-\bar{X}^{(p)}\right)+\sum_{j\neq p}^{s}\alpha_j\left(X_i^{(j)}-\bar{X}^{(j)}\right)\right] + G_i\delta_i\beta^T X_i \\ &\quad - G_i H_0(Y_i)e^{\beta^T\bar{X}}\left[1+\beta_p\left(X_i^{(p)}-\bar{X}^{(p)}\right)+\sum_{j\neq p}^{s}\beta_j\left(X_i^{(j)}-\bar{X}^{(j)}\right)\right] + \text{constant} \\ &= \frac{-1}{2\sigma^2}\left(X_i^{(p)} - \theta^T X_i^{(-p)} - \theta_0\right)^2 + \\ &\quad \left[G_i\alpha_p - \frac{e^{\alpha^T\bar{X}+\alpha_0}}{1+e^{\alpha^T\bar{X}+\alpha_0}}\alpha_p + G_i\delta_i\beta_p - G_i H_0(Y_i)e^{\beta^T\bar{X}}\beta_p\right]X_i^{(p)} + \text{constant}\end{aligned} \tag{3}$$

If we complete the square on (3), we see that the mean of this normal distribution will be a linear combination of $X_i^{(-p)}$, G_i, $G_i \times \delta_i$, and $G_i \times H_0(Y_i)$. A second order Taylor approximation of $e^{\alpha^T X_i}$ and $e^{\beta^T X_i}$ will also give the interaction $G_i \times H_0(Y_i) \times X_i^{(-p)}$. This suggests that when $X^{(p)}$ is normal and the assumptions are satisfied, we can approximate the exact distribution $f(X_i^{(p)}|G_i,\delta_i,Y_i,X_i^{(-p)})$ using a linear regression model with $X_i^{(-p)}$, G_i, $G_i \times \delta_i$, $G_i \times H_0(Y_i)$ and perhaps $G_i \times H_0(Y_i) \times X_i^{(-p)}$ as predictors. In the appendix, we include a similar derivation for an approximate imputation model when $X^{(p)}$ is binary, and $X^{(p)}$ has a logistic relation to $X^{(-p)}$. In the binary case, we approximate the exact distribution using a logistic regression model with $X_i^{(-p)}$, G_i, $G_i \times \delta_i$, $G_i \times H_0(Y_i)$, and $G_i \times H_0(Y_i) \times X_i^{(-p)}$ as covariates. We will call this imputation approach the 'Approximate Cure' approach.

The approximate imputation models implied by (3) and (A.2) explicitly depend on $H_0(t)$ and G_i. To use the derived approximate distributions for covariate imputation, we estimate $H_0(t)$ and impute G_i as part of the chained equations algorithm as we did in Section 3.2. In contrast, the imputation approaches discussed in Sections 3.3.1– 3.3.3 do not require us to impute G_i.

The final interaction term in the imputation models implied by (3) and (A.2) may has many parameters if X_i consists of many covariates, so that term may have to be dropped for settings with many covariates. Also, it may be that the imputed G_i and $G_i \times \delta_i$ are highly correlated, so one may need to only use G_i because of collinearity issues.

In order to impute partially observed covariates using these approximations, we can perform a modification of the Exact Cure algorithm proposed in Section 3.2. We can impute using MICE in R by iterating the following steps: Step (1), estimate $H_0(t)$ as in Section 3.2; Step (2), impute cure status as in Section 3.2; and Step (3), impute each partially observed covariate $X^{(p)}$ sequentially using an appropriate elementary imputation method in MICE (e.g., mice.impute.logreg() for binary covariates) with predictors $X_i^{(-p)}$, G_i, $G_i \times \delta_i$, $G_i \times \hat{H}_0(Y_i)$ and perhaps $G_i \times \hat{H}_0(Y_i) \times X_i^{(-p)}$ [19].

A natural alternative to the proposed Approximate Cure approach is to first impute G and then impute covariates separately for the $G = 1$ and $G = 0$ groups. We could then apply imputation approaches for survival data without a cure fraction (such as the White and Royston method) for imputing covariates in the $G = 1$ group. In simulations (not shown), this approach resulted in similar bias and inflated variances compared with the Approximate Cure approach.

3.3.5. A modification: event time imputation. Because the observed event/censoring time $Y = \min(T, C)$ is a mixture of two underlying random variables, it may not be very intuitive to include Y as a predictor in standard regression models for imputing missing covariates. Instead, we may wish to include the true event time, T, which is not fully observed. We can treat T as another partially observed variable and

impute values of T for censored individuals within the chained equations algorithm used to impute missing covariates. This modification can conceptually be applied to any of the imputation approaches we have discussed.

In the cure setting, T is defined as infinity for cured individuals and is an event time for non-cured individuals. Although cure status is not known for censored individuals, if we also impute G as part of the chained equations imputation algorithm, then we can impute values of T for the non-cured, censored subjects using an assumed truncated distribution $f(t|t > C, G = 1, X)$. We can modify the Exact and Approximate Cure imputation algorithms by adding a step to the chained equations imputation algorithm to impute T_i for censored individuals who have $G_i = 1$ at iteration k. Then, we replace (Y_i, δ_i) in the subsequent imputation models for the partially observed covariates with the imputed (T_i, G_i). In several simulations (not shown), however, T imputation does not appear to improve the performance of the Exact Cure and Approximate Cure imputation algorithms.

We are interested to see how some simple covariate imputation approaches for survival data without a cure fraction are impacted by first imputing T and then substituting (Y, δ) by $(T, 1)$ in the covariate imputation models. We consider both the logY and Outcome Binning approaches. For the Outcome Binning approach, we use octiles to define bins of T among all subjects. In these two approaches, cure status is not known or imputed for censored individuals, and so we cannot impute censored T using the truncated distribution $f(t|t > C, X, G = 1)$. Instead, we impute the event time T using the truncated distribution $f(t|t > C, X)$, which we assume has a proportional hazards structure with a Weibull baseline.

We use a CPH model for the hazard of an event in the total study population. The survival function of the truncated distribution $f(t|t > C_i, X_i)$ of T_i is in the form $S_{TRUNC}(t|X_i) = e^{-[H_0(t)-H_0(C_i)]e^{\beta^T X_i}}, t > C_i$. To impute T_i for a censored individual, we can first generate U_i from a Uniform (0,1) distribution. We can then draw T_i using the relation $T_i = H_0^{-1}\left(-\log(U_i)e^{-\beta^T X_i} + H_0(C_i)\right)$. This requires us to draw β and estimate $H_0(t)$. If we assume the failure time is Weibull such that $S(t|X_i) = e^{-\lambda t^\eta e^{\beta^T X_i}}$, then we can generate T_i as $T_i = \left(\frac{-\log(U_i)e^{-\beta^T X_i} + \lambda C_i^\eta}{\lambda}\right)^{1/\eta}$ after drawing values for β, λ, and η. Within the chained equations algorithm, we generate a T_i value for all censored subjects at each iteration. We can obtain draws of β, λ, and η by first fitting a Weibull regression model to the entire study population using the most recent imputed X and then drawing β, λ, and η from a multivariate normal distribution with mean and covariance estimated by the Weibull fit.

We note that in the CPH cure model setting, the truncated distribution $f(t|t > C, X)$ is incorrectly specified, and it may seem unintuitive to use this misspecified model to impute event times. However, event time imputation has been used in the non-cure survival setting, and an analyst might naively try to apply the same approach to survival data with a cure fraction [24]. We want to see whether this approach improves or worsens the performance of imputation approaches for survival data without a cured fraction when applied in the cure setting.

4. Simulations

In this section, we present results from a simulation study to compare the imputation approaches in terms of bias, relative variance, and coverage rate of confidence intervals for estimating CPH cure model parameters. We also compare with CC analysis and analysis of the full data without any covariate missingness.

4.1. Simulation details

We create 500 simulated datasets of 500 observations each. For each dataset, we simulate multivariate normal covariates $X = (X_1, X_2)$ with zero means, unit variances, and a correlation of 0.5. We then simulate cure status using the relation $\text{logit}\left(P(G_i = 1|X_{i,1}, X_{i,2})\right) = 0.5 + 0.5X_{i,1} + 0.5X_{i,2}$, leading to an average cure rate of 40%. For the non-cured group, we simulate a survival time T_i. We model the event hazard in the non-cured group as $h(t) = h_0(t)e^{0.5X_1+0.5X_2}$ with $h_0(t) = 0.002$. We then generate censoring times $C_i \sim U(250, 4500)$ and define $Y_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leqslant C_i)$.

We impose ~50–55% missingness in X_2 using three models: (1) missing completely at random (MCAR) with $P(X_2 \text{ missing}|X_1, \delta, Y) = 0.5$, (2) MAR with $\text{logit}(P(X_2 \text{ missing}|X_1, \delta, Y)) = X_1$, and (3) MAR with $\text{logit}(P(X_2 \text{ missing}|X_1, \delta, Y)) = 0.3 - 0.4\delta - 0.5X_1\delta$. While this final missingness mechanism may seem implausible, it could be induced when missingness depends on an unobserved variable U that is independently related to T.

We note that we impose missingness in only a single covariate rather than many covariates (the typical setting where FCS is applied). However, we are mainly interested in investigating various strategies for modeling the univariate conditional distribution for one partially observed covariate. As such, we can compare the imputation approaches by imposing missingness in only one covariate. Similar results can be seen when we apply the imputation approaches with missingness in multiple covariates (Supporting Information). We also consider the setting with many partially observed covariates in our head and neck cancer example.

We perform multiple imputation of X_2 using methods described in this paper. For each simulation and method, we produce 10 imputed datasets. We then fit a CPH cure model to each imputed dataset (ignoring imputed cure status) and use Rubin's Rules to obtain a single set of estimates for each simulation [20]. We then compute bias, relative variance (compared with analyzing the full data with no covariate missingness), and coverage in estimating model parameters across 500 simulations for each method. Alternatively, for imputation approaches that result in imputed values for G, we could have performed our final analysis by fitting Cox and logistic regressions given the imputed G. In simulations (not shown), this approach resulted in a slight increase in efficiency for estimating the intercept for the logistic part of the model, but it also resulted in some increases in bias for the approaches using approximated distributions for imputation.

We use 100 iterations for each imputation algorithm except Exact Cure, for which we use 1500 because of the slower convergence of the Metropolis–Hastings algorithms. When fitting the cure models to each imputed dataset, we use 100 iterations of the EM algorithm and use 100 bootstrap samples of the imputed dataset to estimate variances.

Computational time is shortest for the Outcome Bins and logY approaches, followed closely by the T imputation methods. The Approximate Cure approach takes about four times as long as the Outcome Bins method to run and about two times as long as the White and Royston method. The Exact Cure approach takes at least 10 times as long as the Approximate Cure approach to run.

4.2. Simulation results

Table I shows simulation results under three different missingness mechanisms for X_2. Under missingness models (1) and (2), CC analysis is essentially unbiased. However, in model (3), CC analysis results in biased estimates, particularly in estimating parameters for the logistic part of the mixture cure model. In all missingness settings shown, the imputation methods have little bias in estimating α_0, α_1, and β_1, the logistic model intercept and the parameters associated with X_1.

In all three missingness settings, the logY, White and Royston, Outcome Binning, T imputation, and Approximate Cure (w/o extra interaction) approaches result in similar or larger bias than CC analysis in estimating α_2, the logistic parameter for X_2. For all three missingness models, the imputation approaches using T imputation result in larger α_2 bias than their counterparts without T imputation. The Approximate Cure approach with the interaction term and the Exact Cure approach produce comparably low bias in estimating α_2.

All imputation methods except the Exact Cure approach result in biased estimates for β_2, the failure time model parameter associated with X_2. Among the biased imputation methods, however, the Approximate Cure approach including the extra interaction term consistently results in the smallest β_2 bias. The logT approach produces smaller β_2 bias than the logY approach. Outcome Binning results in similar β_2 bias with and without the T imputation.

All imputation methods result in smaller empirical variance (so larger relative variance) in estimating α_0, α_1, and β_1 compared with CC analysis in all three simulation settings. Some reduction in the variance in estimating β_2 can also be seen, suggesting that we can still gain some information about the effect of X_2 by including information from subjects with missing X_2. Coverage rates for α_0, α_1, and β_1 are similar for all imputation methods in all three simulation settings. CC coverage of 95% confidence intervals for α_0 and α_1 under missingness model (3) is far below 0.95%. Reductions in coverage for some imputation approaches can be seen for α_2 and β_2. Undercoverage is mainly because of increased bias. The Exact Cure approach and the Approximate Cure approach with the extra interaction term tend to produce higher coverage rates in estimating β_2 compared with the other imputation methods.

In all three sets of simulations, we see large reductions in the Approximate Cure approach's corresponding biases by adding the extra interaction term. Although not shown, we do not see corresponding decreases in bias by adding a $\hat{H}_0(Y_i) : X^{(-p)}$ interaction term to the White and Royston approach [15].

Table I. Bias, relative variance, and coverage of cure model estimates across 500 simulations.

Method	α_0 Bias (RV) CI	α_1 Bias (RV) CI	α_2 Bias (RV) CI	β_1 Bias (RV) CI	β_2 Bias (RV) CI
Full data	−0.01 (1.00) 0.93	0.02 (1.00) 0.93	0.02 (1.00) 0.94	−0.01 (1.00) 0.95	0.00 (1.00) 0.95
Missingness Model 1: MCAR missingness in X_2					
Exact cure	0.00 (0.83) 0.94	0.01 (0.75) 0.92	0.03 (0.48) 0.94	−0.01 (0.82) 0.94	0.00 (0.48) 0.95
Approximations					
Non-cure w/ (Y, δ)					
logY	0.00 (0.79) 0.94	0.00 (0.74) 0.91	0.08 (0.47) 0.92	0.01 (0.85) 0.95	−0.14 (0.71) 0.78
White and Royston	0.00 (0.81) 0.94	0.00 (0.73) 0.93	0.07 (0.47) 0.92	0.00 (0.82) 0.95	−0.13 (0.76) 0.81
Binning by (Y, δ)	0.00 (0.80) 0.94	0.01 (0.75) 0.93	0.04 (0.48) 0.93	0.00 (0.83) 0.96	−0.11 (0.66) 0.87
Non-cure w/ T					
logT	0.00 (0.80) 0.94	−0.02 (0.79) 0.93	0.14 (0.55) 0.89	0.01 (0.94) 0.96	−0.12 (0.90) 0.86
Binning by T	0.00 (0.81) 0.94	0.00 (0.78) 0.93	0.09 (0.53) 0.92	0.00 (0.86) 0.95	−0.10 (0.71) 0.89
Cure w/ (G, Y, δ)					
Approx cure	0.00 (0.85) 0.94	0.01 (0.75) 0.93	0.05 (0.50) 0.94	0.00 (0.81) 0.95	−0.13 (0.82) 0.82
Approx + Int*	0.00 (0.85) 0.93	0.02 (0.78) 0.92	0.02 (0.47) 0.93	0.00 (0.91) 0.95	−0.07 (0.75) 0.93
Complete case	−0.01 (0.48) 0.94	0.03 (0.52) 0.96	0.03 (0.49) 0.94	0.00 (0.52) 0.97	0.00 (0.46) 0.95
Missingness Model 2: MAR missingness in X_2 dependent on X_1					
Exact cure	0.00 (0.84) 0.95	0.01 (0.81) 0.94	0.04 (0.47) 0.93	0.00 (0.79) 0.95	−0.01 (0.34) 0.92
Approximations					
Non-cure w/ (Y, δ)					
logY	0.00 (0.82) 0.94	0.01 (0.82) 0.95	0.13 (0.47) 0.91	0.02 (0.80) 0.95	−0.20 (0.63) 0.62
White and Royston	0.00 (0.79) 0.95	−0.02 (0.79) 0.95	0.14 (0.46) 0.89	0.02 (0.77) 0.96	−0.19 (0.63) 0.65
Binning by (Y, δ)	0.00 (0.83) 0.95	0.00 (0.80) 0.94	0.10 (0.51) 0.92	0.01 (0.78) 0.95	−0.16 (0.54) 0.73
Non-cure w/ T					
logT	0.01 (0.83) 0.94	−0.01 (0.85) 0.94	0.15 (0.61) 0.88	0.02 (0.85) 0.95	−0.16 (0.71) 0.73
Binning by T	0.00 (0.84) 0.94	0.00 (0.82) 0.95	0.11 (0.58) 0.93	0.01 (0.83) 0.95	−0.15 (0.53) 0.76
Cure w/ (G, Y, δ)					
Approx cure	0.00 (0.84) 0.94	−0.01 (0.76) 0.94	0.12 (0.49) 0.90	0.02 (0.71) 0.94	−0.20 (0.67) 0.61
Approx + Int*	0.00 (0.89) 0.95	0.01 (0.82) 0.94	0.05 (0.48) 0.94	0.00 (0.78) 0.94	−0.12 (0.65) 0.86
Complete case	0.00 (0.41) 0.95	0.04 (0.43) 0.94	0.05 (0.50) 0.95	−0.02 (0.31) 0.95	−0.02 (0.33) 0.91
Missingness Model 3: MAR missingness in X_2 dependent on X_1, δ					
Exact cure	0.00 (0.86) 0.94	0.01 (0.77) 0.93	0.03 (0.44) 0.94	−0.01 (0.82) 0.95	0.00 (0.60) 0.95
Approximations					
Non-cure w/ (Y, δ)					
logY	0.00 (0.81) 0.93	0.00 (0.77) 0.94	0.07 (0.42) 0.93	0.00 (0.88) 0.95	−0.11 (0.88) 0.89
White and Royston	0.00 (0.83) 0.94	0.00 (0.79) 0.93	0.06 (0.44) 0.94	0.00 (0.87) 0.96	−0.09 (0.87) 0.90
Binning by (Y, δ)	0.00 (0.86) 0.94	0.02 (0.79) 0.94	0.02 (0.44) 0.94	0.00 (0.81) 0.96	−0.08 (0.71) 0.92
Non-cure w/ T					
logT	0.01 (0.83) 0.94	−0.03 (0.81) 0.92	0.16 (0.52) 0.88	0.01 (0.94) 0.96	−0.09 (0.99) 0.92
Binning by T	0.00 (0.84) 0.94	0.00 (0.80) 0.94	0.09 (0.47) 0.92	0.00 (0.86) 0.96	−0.07 (0.78) 0.94
Cure w/ (G, Y, δ)					
Approx cure	0.00 (0.87) 0.93	0.02 (0.78) 0.94	0.02 (0.44) 0.95	−0.01 (0.81) 0.96	−0.08 (0.94) 0.92
Approx + Int*	0.00 (0.88) 0.93	0.02 (0.82) 0.94	0.03 (0.44) 0.94	0.00 (0.89) 0.96	−0.05 (0.93) 0.95
Complete Case	0.18 (0.39) 0.83	0.29 (0.41) 0.77	0.03 (0.43) 0.96	0.00 (0.54) 0.95	0.00 (0.57) 0.95

*Includes $\hat{H}_0(Y) : G : X_1$ interaction in imputation model.
CI indicates empirical coverage of 95% confidence intervals and RV indicates relative variance.

We also see that the Exact Cure imputation approach far outperforms all other imputation algorithms in terms of bias, and among the biased imputation approaches, the Approximate Cure approach with the interaction term is generally the best performer. In all three sets of simulations, the non-cure imputation approaches that involve T imputation tend to have worse coverage or bias properties than the corresponding approaches without T imputation. Finally, we see that among the approaches that do not take the cure fraction into account (Outcome Binning, logY, White and Royston, and logT), Outcome Binning without T imputation tends to produce the smallest bias overall across the three simulation settings.

5. Head and neck cancer example

We consider data from a cohort study of time to cancer recurrence in $N = 1226$ patients with head and neck squamous cell carcinoma (HNSCC). This study was conducted by the University of Michigan's Head and Neck Specialized Program of Research Excellence (SPORE) and included consenting patients treated for HNSCC at the University of Michigan Cancer Center between November 2003 and July 2013. Details regarding the cohort study can be found in Duffy *et al*. and Virani *et al*. [25, 26]. Data on newly-diagnosed patients were collected from the time of diagnosis, and patients were then followed for cancer recurrence after the start of treatment. A patient is considered to have recurred if cancer becomes detectable. Personal and disease-related characteristics including age, cancer stage, cancer site, comorbidities, cigarette use, alcohol use, gender, and body mass index (BMI) were collected at the time of diagnosis and are reported in Table II.

Of the 1226 patients in the study, 374 (30.5%) experienced a cancer recurrence. Of these, 149 (39.8%) had detectable cancer toward the end of their planned treatment. These patients are called 'persistent' and are given a recurrence time of 1 day as exact recurrence times are unavailable for these subjects. Patients were followed for a median time of 36.6 months. Of the observed recurrences, 360 (96.2%) occurred within 36 months. Few patients had recurrences after 36 months, and the estimated survival curve had a plateau in the later half of the study (∼36–60 months). For HNSCC, it is well established that patients can be cured [27]. This provides some evidence that these data may follow a cure structure.

Based on biological knowledge of HNSCC recurrence and empirical evidence in the data, we assume that a subset of the study cohort had been cured of disease by treatment, and we fit a mixture cure model. We assume a CPH model for the hazard of cancer recurrence in the non-cured group, and we model probability of being cured of the primary HNSCC after treatment using a logistic regression. In particular, the first component is a model for time until cancer becomes detectable in the non-cured group. We include persistent patients in our analysis as persistence was defined subjectively and roughly corresponded to whether there were early signs that the cancer was present. Because persistence is an outcome of the treatment that was unobserved at baseline, these patients were included in the analysis. We fit a CPH cure model to the CC data using age at diagnosis, cancer stage, cigarette use, human papillomavirus (HPV) status, comorbidities, and cancer site as predictors in both parts of the mixture cure model. Results of this model fit are shown in Table III.

Table II. Characteristics of $N = 1226$ study patients at HNSCC Diagnosis.

Characteristic	*N* (%) or Mean (SD)	Missing *N* (%)	Characteristic	*N* (%) or Mean (SD)	Missing *N* (%)
		Model variables			
Age at diagnosis	59.5 (11.7)		Comorbidities		1 (0.01)
Cancer stage		0 (0)	None	343 (27.9)	
I/Cis	162 (13.2)		Mild	535 (43.6)	
II	123 (10.0)		Moderate	239 (19.4)	
III	181 (14.7)		Severe	108 (8.8)	
IV	760 (61.9)		Cancer site		0 (0)
Cigarette use		0 (0)	Larynx	245 (19.9)	
Never	285 (23.2)		Hypopharynx	53 (4.3)	
Current	559 (45.5)		Oral cavity	413 (33.6)	
Former	382 (31.1)		Oropharynx	515 (42.0)	
HPV status		685 (55.8)			
Negative	320 (26.1)				
Positive	221 (18.0)				
		Auxiliary variables			
Gender		0 (0)	Enrollment year		0 (0)
Female	315 (25.6)		2003–2008	559 (45.5)	
Male	911 (74.3)		2009–2011	363 (29.6)	
Alcohol use		1 (0.01)	2012–2013	304 (24.7)	
Never	115 (9.3)		No. sexual partners	16.8 (53.4)	765 (62.3)
Current	300 (24.3)		BMI	26.9 (5.9)	6 (0.4)
Former	810 (66.0)				

Table III. Cox proportional hazards cure model of time-to-HNSCC recurrence.

Patient Characteristic	Complete case analysis, $N = 540$		Approx cure + int*, $N = 1226$		White and Royston, $N = 1226$	
	Logistic OR, 95% CI	Failure time HR, 95% CI	Logistic OR, 95% CI	Failure time HR, 95% CI	Logistic OR, 95% CI	Failure time HR, 95% CI
Age at diagnosis						
10 Year ↑	1.07 (0.91, 1.26)	1.23 (1.02, 1.45)†	1.14 (1.00, 1.31)†	1.08 (0.98, 3.95)	1.14 (0.99, 1.31)	1.08 (0.98, 1.18)
Cancer stage						
I/Cis (ref)						
II	0.94 (0.31, 2.88)	2.17 (0.51, 9.23)	1.25 (0.57, 2.74)	1.67 (0.70, 3.95)	1.26 (0.56, 2.84)	1.68 (0.66, 4.28)
III	2.25 (0.84, 6.00)	2.91 (0.72, 11.6)	2.36 (1.18, 4.72)†	2.42 (1.22, 4.79)†	2.31 (1.19, 4.47)†	2.42 (1.13, 5.19)†
IV	2.42 (1.11, 5.31)†	2.77 (0.68, 11.1)	3.32 (1.74, 6.33)†	2.76 (1.48, 5.16)†	3.25 (1.84, 5.75)†	2.78 (1.39, 5.59)†
Cigarette use						
Never (ref)						
Current	1.03 (0.57, 1.89)	0.63 (0.34, 1.14)	1.46 (0.97, 2.18)	0.98 (0.70, 1.38)	1.49 (1.00, 2.21)†	0.99 (0.72, 1.35)
Former	1.09 (0.63, 1.87)	0.56 (0.35, 0.90)†	1.27 (0.85, 1.90)	0.94 (0.66, 1.33)	1.28 (0.84, 1.93)	0.95 (0.69, 1.32)
HPV status						
Negative (ref)						
Positive	0.43 (0.21, 0.87)†	0.80 (0.35, 1.82)	0.34 (0.19, 0.58)†	0.91 (0.55, 1.48)	0.38 (0.19, 0.76)†	0.82 (0.52, 1.28)
Comorbidities						
None (ref)						
Mild	1.14 (0.66, 1.97)	0.93 (0.48, 1.81)	1.14 (0.77, 1.69)	0.89 (0.65, 1.23)	1.14 (0.80, 1.62)	0.89 (0.65, 1.21)
Moderate	1.32 (0.65, 2.68)	1.47 (0.72, 2.98)	1.66 (1.08, 2.56)†	1.10 (0.75, 1.61)	1.68 (1.12, 2.53)†	1.09 (0.74, 1.60)
Severe	1.70 (0.73, 3.92)	0.79 (0.24, 2.60)	1.94 (1.10, 3.43)†	1.07 (0.63, 1.80)	1.96 (1.09, 3.52)†	1.05 (0.62, 1.80)
Cancer site						
Larynx (ref)						
Hypopharynx	7.90 (0.00, Inf.)	2.42 (0.88, 6.64)	1.93 (0.88, 4.22)	1.43 (0.77, 2.67)	1.91 (0.88, 4.16)	1.46 (0.76, 2.80)
Oral cavity	1.58 (0.83, 3.00)	1.33 (0.61, 2.89)	1.24 (0.81, 1.90)	1.33 (0.90, 1.97)	1.24 (0.81, 1.90)	1.34 (0.92, 1.95)
Oropharynx	1.51 (0.66, 3.44)	0.93 (0.39, 2.18)	1.68 (0.94, 3.02)	1.02 (0.62, 1.68)	1.57 (0.84, 2.94)	1.11 (0.69, 1.78)

*Includes $\hat{H}_0(Y) : G : X^{(-p)}$ interaction in imputation model
†Significant at $p = 0.05$

In the study of HNSCC, the association between HPV status and cancer recurrence is of particular interest. However, HPV status was only obtained for 541 (44.1%) of the patients. Investigation into the missingness of HPV status (not shown) suggests that HPV missingness is associated with diagnosis date and therefore censoring time. However, assuming censoring is independent of HPV status, we can still assume HPV status is MAR [28]. We want to impute HPV status using approaches discussed and then compare results from corresponding CPH cure model estimates between imputation approaches and to CC analysis.

We performed multiple imputation of HPV status (55.8% missing) and comorbidities (0.01% missing) using both the Approximate Cure approach with the extra interaction term and the White and Royston approach. We did not use the Exact Cure approach as we have many partially observed covariates, and when we have many covariates to impute, the Exact Cure approach becomes increasingly computationally intensive. HPV status is known to be associated with factors such as gender, smoking, alcohol use, and number of sexual partners. HPV also has a much higher prevalence for oropharyngeal cancers compared with other types of head and neck cancer. We observe that HPV status is associated with calendar time and therefore year of study enrollment. As these variables are known to be associated with HPV status, they may help us to obtain better imputations of HPV. Therefore, we use all factors in Table II as predictors for the various imputation models, requiring us to also impute BMI, number of sexual partners, and alcohol use as part of the chained equations algorithm. We note that sexual partners has a large amount of missingness (62.3%), but we include it in the imputation algorithm because of its strong association with HPV status. Number of sexual partners is observed for 198 (28.9%) of the subjects with missing HPV status. Year of study enrollment was categorized into three intervals reflecting different rates of HPV missingness. Greater effort was made to obtain HPV status for subjects enrolled after 2008, and some samples obtained in 2012 and 2013 have not yet been tested. Some of the Table II variables are not included in the final cure model analysis as cure models become increasingly unstable with a large amount of predictors. We therefore implicitly assume that the predictors not included in the final model are not independent predictors of the outcome. In order to satisfy the assumptions made in the derivation of the Approximate Cure approach, we assume that censoring of recurrence time (including death from other causes) does not depend on the partially observed variables and in particular HPV status and number of sexual partners. We impute categorical covariates using polytomous regression in MICE [19]. Number of sexual partners is imputed using predictive mean matching on the log-scale. We produced 20 imputed datasets for each approach.

Table III shows the CPH cure model results for two imputation algorithms and CC analysis. Point estimates and confidence intervals are very similar between the two imputation approaches. Based on the simulation results, we may expect the biggest difference between the two approaches to be the bias in estimating parameters for HPV status. For this dataset, however, the estimates for the parameters corresponding to HPV status are very similar between the two imputation approaches. When we apply other imputation approaches discussed in this paper to these data (not shown), we see similar results.

Differences can be seen between the model fits from imputation and from CC analysis. Confidence intervals tend to be narrower for the imputation approaches than for CC analysis. Point estimates tend to be somewhat similar with some exceptions. The most notable difference between the imputation and complete case fits is in the estimates for the cigarette use variable. Point estimates from the imputation approaches suggest that cigarette use may be associated with a decrease in the probability of being cured, but it is not associated with the hazard of recurrence. In contrast, the complete case analysis suggests that cigarette use is associated with a decreased hazard of recurrence in the non-cured group, but it is not associated with cure status. Additionally, the confidence intervals for some cigarette use parameters from the imputation approaches do not include the complete case point estimates. The complete case fit shows some signs of model instability.

Point estimates for HPV status parameters are similar between the complete case and imputation approaches, but the confidence intervals are smaller in the imputation model fits. This suggests that some additional information about HPV status is obtained by including information from the patients with missing HPV status.

6. Discussion

In this paper, we have explored approaches for imputing missing covariates in the CPH cure model setting. We considered multiple imputation using FCS, an approach in which we impute partially observed covariates by drawing from their conditional distributions.

We derived the exact conditional distribution and suggested a sampling scheme for imputing normal and Bernoulli covariates in the CPH cure model setting. We also proposed several approximations to the exact distribution that are simpler and more convenient to use for imputation. Our approach can be generalized to impute covariates with different distributions. We compared the performance of our proposed imputation approaches with existing imputation methods for survival data without a cure fraction.

A simulation study demonstrates that all imputation methods considered can substantially increase precision in estimating many CPH cure model parameters compared with complete case analysis. Imputation can produce smaller variances for estimating parameters corresponding to fully observed variables compared with complete case analysis. Some variance reduction may also be seen in estimating parameters associated with the imputed variables. The Exact Cure imputation approach outperformed all other imputation approaches in terms of bias. In our simulations, all other imputation approaches tended to have some bias in estimating at least one of the parameters associated with the imputed variable/s. Among the biased imputation approaches, the Approximate Cure approach with the interaction term was the best performer. Among the approaches that do not account for the cure fraction, Outcome Binning tended to have the best performance across the three simulation settings. The approaches in which the event time is imputed without accounting for the cure structure of the data did not perform well in the cure setting and are not recommended. In the head and neck cancer example, little difference could be seen between the imputation approaches, but many differences were present between imputation and complete case analysis.

While imputation using the exact conditional distribution is a clear frontrunner in terms of bias, it is typically more difficult to implement and takes much longer to run than other methods because of the many required Metropolis–Hastings draws. These issues become even more pronounced when there is missingness in multiple covariates. If one is willing to allow some bias in estimating some model parameters (particularly those associated with the imputed variables), then the Approximate Cure imputation approach with the interaction term may be preferred. For example, if we are only adjusting for an imputed variable as a possible confounder, then adding some bias in estimating its parameters in exchange for computational simplicity may be acceptable. If we desire an even simpler imputation scheme and do not want to impute cure status, we may still be able to obtain some bias reduction by using Outcome Binning without the event time imputation rather than other existing imputation approaches for survival data without a cure fraction.

We compare imputation approaches in terms of performance in estimating CPH cure model parameters, and most of the imputation approaches proposed are compatible with and directly motivated by the final modeling strategy. If we change the modeling strategy (for example, if we want to fit an accelerated failure time model with a cure fraction), then the imputation approach may need to be adapted and the comparative performance of the approaches may change. Additionally, although simulations suggest there is a difference between imputation approaches, there may not always be a large practical difference when applied to particular datasets as seen with the head and neck cancer data. The presented simulations are limited to a setting with normal and binary covariates with linear covariate effects in the logistic and failure time models. When imputing covariates with other distributions (e.g., ordered categorical), the comparative performance of the imputation approaches may be different. Also, if the failure time or logistic models include interactions/non-linear effects of the partially observed covariates, the difference between the Exact Cure method and the approximated methods would be expected to be even more pronounced than in the linear effects case considered here [14].

We note that $H_0(t)$ in the CPH model is really an infinite-dimensional parameter, and we do not directly incorporate this uncertainty into the estimation procedure. Additionally, we only consider multiple imputation using FCS. FCS is convenient to use for imputation as it does not require us to explicitly specify the joint distribution of the covariates. However, in the case of multiple imputed variables, the assumed distributions for each partially observed $X^{(p)}|X^{(-p)}$ are not guaranteed to be compatible and form a valid joint distribution. In some cases, this could lead to problems (e.g., bias) when estimating parameters in the final model fitting [14]. Several authors have provided conditions in which FCS is equivalent to joint model imputation and converges to the desired sampling distribution [29, 30].

Appendix A: imputation for binary covariates

We will derive an approximate imputation model for imputing binary covariates. Suppose $X^{(p)} \sim$ Bernoulli(t) where $t = expit(\theta^T X^{(-p)} + \theta_0)$. Using the complete data likelihood for the CPH cure model,

we have

$$\begin{aligned}
&\operatorname{logit}\left(P\left(X_i^{(p)}=1|G_i,\delta_i,Y_i,X_i^{(-p)}\right)\right)=\log\left(\frac{L(\alpha,\alpha_0,\beta,\theta,\theta_0,\gamma)|_{X_i^{(p)}=1}}{L(\alpha,\alpha_0,\beta,\theta,\theta_0,\gamma)|_{X_i^{(p)}=0}}\right)\\
&=\log\left(\frac{\left\{\left(e^{\beta_p}\right)^{\delta_i}e^{-H_0(Y_i)e^{\beta_p+\sum_{j\neq p}^s X_i^{(j)}\beta_j}}\frac{e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j+\alpha_p}}{1+e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j+\alpha_p}}\right\}^{G_i}\times\left\{\frac{1}{1+e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j+\alpha_p}}\right\}^{1-G_i}e^{\theta^T X^{(-p)}+\theta_0}}{e^{-H_0(Y_i)e^{\sum_{j\neq p}^s X_i^{(j)}\beta_j}}\times\left\{\frac{e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j}}{1+e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j}}\right\}^{G_i}\times\left\{\frac{1}{1+e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j}}\right\}^{1-G_i}}\right)\\
&=\theta_0+\theta^T X_i^{(-p)}+G_i\delta_i\beta_p-G_iH_0(Y_i)\left(e^{\beta_p}-1\right)e^{\sum_{j\neq p}^s X_i^{(j)}\beta_j}+\alpha_pG_i+\log\left(1+e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j}\right)\\
&\quad-\log\left(1+e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j+\alpha_p}\right)
\end{aligned}\tag{A.1}$$

This relation gives the form for the exact conditional distribution, which we can use to impute a partially observed, binary $X^{(p)}$. Now, we attempt to find a simpler approximated model. We use a similar approach as in the normal derivation. Assuming $\operatorname{Var}(\alpha^T X_i)$ and $\operatorname{Var}(\beta^T X_i)$ are small, we approximate the above by

$$\begin{aligned}
&\operatorname{logit}\left(P\left(X_i^{(p)}=1|G_i,\delta_i,Y_i,X_i^{(-p)}\right)\right)\approx\theta_0+\theta^T X_i^{(-p)}+G_i\delta_i\beta_p-G_iH_0(Y_i)\left(e^{\beta_p}-1\right)e^{\sum_{j\neq p}^s X_i^{(j)}\beta_j}+\alpha_pG_i\\
&\qquad+\frac{e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j}}{1+e^{\alpha_0+\sum_{j\neq p}^s \bar{X}^{(j)}\alpha_j}}-\frac{e^{\alpha_0+\sum_{j\neq p}^s X_i^{(j)}\alpha_j+\alpha_p}}{1+e^{\alpha_0+\alpha_p+\sum_{j\neq p}^s \bar{X}^{(j)}\alpha_j}}+\text{constant}\\
&\approx\theta_0+\theta^T X_i^{(-p)}+G_i\delta_i\beta_p-G_iH_0(Y_i)\left(e^{\beta_p}-1\right)e^{\sum_{j\neq p}^s \bar{X}^{(j)}\beta_j}\left[1+\sum_{j\neq p}^s\beta_j\left(X_i^{(j)}-\bar{X}^{(j)}\right)\right]\\
&\qquad+\alpha_pG_i+\frac{e^{\alpha_0+\sum_{j\neq p}^s \bar{X}^{(j)}\alpha_j}}{1+e^{\alpha_0+\sum_{j\neq p}^s \bar{X}^{(j)}\alpha_j}}\left[1+\sum_{j\neq p}^s\alpha_j\left(X_i^{(j)}-\bar{X}^{(j)}\right)\right]\\
&\qquad-\frac{e^{\alpha_0+\sum_{j\neq p}^s \bar{X}^{(j)}\alpha_j+\alpha_p}}{1+e^{\alpha_0+\alpha_p+\sum_{j\neq p}^s \bar{X}^{(j)}\alpha_j}}\left[1+\sum_{j\neq p}^s\alpha_j\left(X_i^{(j)}-\bar{X}^{(j)}\right)\right]+\text{constant}
\end{aligned}\tag{A.2}$$

This equation is a linear combination of $X_i^{(-p)}$, G_i, $G_i\times\delta_i$, $G_i\times H_0(Y_i)$ and $G_i\times H_0(Y_i)\times X^{(-p)}$. This suggests that we can impute $X_i^{(p)}$ using $X_i^{(-p)}$, G_i, $G_i\times\delta_i$, $G_i\times\hat{H}_0(Y_i)$, $G_i\times\hat{H}_0(Y_i)\times X^{(-p)}$ as predictors in a logistic regression model if we impute G_i for censored subjects and estimate $H_0(Y_i)$ as additional steps in the multiple imputation algorithm.

Acknowledgements

The authors cite the many investigators in the University of Michigan Head and Neck Specialized Program of Research Excellence for their contributions to patient recruitment, specimen collection, study conduct, and encouragement including Emily Bellile, MS, Carol R. Bradford, MD, Thomas E. Carey, PhD, Douglas B. Chepeha, MD, Sonia Duffy, PhD, Avraham Eisbruch, MD, Joseph Helman, DDS, Kelly M. Malloy, MD, Jonathan McHugh, MD, Scott A. McLean, MD, Tamara H. Miller, RN, Jeff Moyer, MD, Lisa Peterson, MPH, Mark E. Prince, MD, Nancy Rogers, RN, Laura Rozek, PhD, Matthew E. Spector, MD, Nancy E. Wallace, RN, Heather Walline, PhD, Brent Ward, DDS, and Francis Worden, MD. We greatly thank our patients and their families who tirelessly participated in our survey and specimen collections.

This research was partially supported by National Institutes of Health grant T32 CA-83654. J. W. B. was supported by a UK Medical Research Council Fellowship (MR/K02180X/1).

References

1. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1982; **38**(4):1041–1046.
2. Yamaguchi K. Accelerated failure-time regression models with a regression model of survival fraction: an application to the analysis of "permanent employment" in Japan. *Journal of the American Statistical Association* 1992; **87**(418):284–292.
3. Lu W, Ying Z. Semiparametric transformation cure models. *Biometrika* 2004; **91**(2):331–343.

4. Kuk AYC, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 1992; **79**(3):531–541.
5. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000; **56**(1):237–243.
6. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000; **56**:227–236.
7. Zhuang D, Schenker N, Taylor JMG, Mosseri V, Dubray B. Analysing the effects of anaemia on local recurrence of head and neck cancer when covariate values are missing. *Statistics in Medicine* 2000; **19**:1237–1249.
8. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society* 1972; **34**(2):187–220.
9. Cho M, Schenker N, Taylor JMG, Zhuang D. Survival analysis with long-term survivors and partially observed covariates. *The Canadian Journal of Statistics* 2001; **29**(3):421–436.
10. Chen MH, Ibrahim JG. Bayesian methods for missing covariates in cure rate models. *Lifetime Data Analysis* 2002; **8**: 117–146.
11. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology* 2006; **59**:1092–1101.
12. van Buuren S, Brands JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation* 2006; **76**(12):1049–1064.
13. Raghunathan TE. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* 2001; **27**(1):85–95.
14. Bartlett JW, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: Accomodating the substantive model. *Statistical Methods in Medical Research* 2014; **24**(4):462–487.
15. White IR, Royston P. Imputing missing covariate values for the cox model. *Statistics in Medicine* 2009; **28**:1982–1998.
16. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 1999; **18**:681–694.
17. Breslow NE. Contribution to the discussion of the paper by DR Cox. *Journal of the Royal Statistical Society* 1972; **24**: 216–217.
18. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* 2nd ed. John Wiley and Sons, Inc: Hoboken, New Jersey, 2002. DOI: 10.1002/9781119013563.
19. van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate imputation by chained equations. *Journal of Statistical Software* 2011; **45**(3):1–67.
20. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons, Inc: Hoboken, New Jersey, 1987, 1–67.
21. Hastings WK. Monte carlo sampling methods using markov chains and their applications. *Biometrika* 1970; **57**:97–109.
22. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 1953; **21**:1087–1092.
23. Sherlock C, Fearnhead P, Roberts GO. The random walk metropolis: linking theory and practice through a case study. *Statistical Science* 2010; **25**(2):170–190.
24. Taylor JMG, Murray S, Hsu CH. Survival estimation and testing via multiple imputation. *Statistics and Probability Letters* 2002; **58**:221–232.
25. Duffy S, Taylor JMG, Terrell J, Islam M, Yuan Z, Fowler K, Wolf G, Teknos T. Il-6 predicts recurrence among head and neck cancer patients. *Cancer* 2008; **113**(4):750–757.
26. Virani S, Bellile E, Bradford CR, Carey TE, Chepeha DB, Colacino JA, Helman JI, McHugh JB, Peterson LA, Sartor MA, Taylor JMG, Walline HM, Wolf GT, Rozek LS. Ndn and cd1a are novel prognostic methylation markers in patients with nead and neck squamous cell carcinomas. *BMC Cancer* 2015; **15**(825):825–838.
27. Taylor JMG. Semiparametric estimation in failure time mixture models. *Biometrics* 1995; **51**(3):899–907.
28. Rathouz PJ. Identifiability assumptions for missing covariates in failure time regression models. *Biostatistics* 2007; **8**(2):345–356.
29. Hughes RA, White IR, Seaman SR, Carpenter JR, Tilling K, Sterne JAC. Joint modeling rationale for chained equations. *BMC Medical Research Methodology* 2014; **14**(28):28–38.
30. Liu J, Gelman A, Hill J, Su YS, Kropko J. On the stationary distribution of iterative imputation. *Biometrika* 2013; **101**(1):155–173.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.