

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



Keogh, RH; Morris, TP (2018) Multiple imputation in Cox regression when there are time-varying effects of covariates. *Statistics in medicine*. ISSN 0277-6715 DOI: <https://doi.org/10.1002/sim.7842>

Downloaded from: <http://researchonline.lshtm.ac.uk/4648590/>

DOI: [10.1002/sim.7842](https://doi.org/10.1002/sim.7842)

Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: <http://creativecommons.org/licenses/by/2.5/>

RESEARCH ARTICLE

Multiple imputation in Cox regression when there are time-varying effects of covariates

Ruth H. Keogh¹  | Tim P. Morris² 

¹Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

²London Hub for Trials Methodology Research, MRC Clinical Trials Unit at UCL, Aviation House, London, UK

Correspondence

Ruth H. Keogh, Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

Email: ruth.keogh@lshtm.ac.uk

Funding information

Medical Research Council Methodology Fellowship, Grant/Award Number: MR/M014827/1; Medical Research Council, Grant/Award Number: MC_UU_12023/21 and MC_UU_12023/29

In Cox regression, it is important to test the proportional hazards assumption and sometimes of interest in itself to study time-varying effects (TVEs) of covariates. TVEs can be investigated with log hazard ratios modelled as a function of time. Missing data on covariates are common and multiple imputation is a popular approach to handling this to avoid the potential bias and efficiency loss resulting from a “complete-case” analysis. Two multiple imputation methods have been proposed for when the substantive model is a Cox proportional hazards regression: an approximate method (Imputing missing covariate values for the Cox model in *Statistics in Medicine* (2009) by White and Royston) and a substantive-model-compatible method (Multiple imputation of covariates by fully conditional specification: accommodating the substantive model in *Statistical Methods in Medical Research* (2015) by Bartlett et al). At present, neither accommodates TVEs of covariates. We extend them to do so for a general form for the TVEs and give specific details for TVEs modelled using restricted cubic splines. Simulation studies assess the performance of the methods under several underlying shapes for TVEs. Our proposed methods give approximately unbiased TVE estimates for binary covariates with missing data, but for continuous covariates, the substantive-model-compatible method performs better. The methods also give approximately correct type I errors in the test for proportional hazards when there is no TVE and gain power to detect TVEs relative to complete-case analysis. Ignoring TVEs at the imputation stage results in biased TVE estimates, incorrect type I errors, and substantial loss of power in detecting TVEs. We also propose a multivariable TVE model selection algorithm. The methods are illustrated using data from the Rotterdam Breast Cancer Study. R code is provided.

KEYWORDS

Cox regression, missing data, multiple imputation, restricted cubic spline, time-varying effect

1 | INTRODUCTION

The setting of this paper is studies of associations between covariates and time-to-event outcomes such as disease diagnosis or death analysed using Cox regression.^{1,2} Missing data in explanatory variables are common and the impacts of ignoring the missing data and performing a “complete-case” analysis on the subset of individuals with no missing data are loss of efficiency and, depending on the missing data mechanism, biased estimates. Multiple imputation (MI) is a widely

.....
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

used approach for handling missing data that involves generating multiple plausible values for the missing data to create multiple imputed data sets. The multiply imputed data sets are each analysed to obtain estimates of interest and corresponding standard errors, which are then combined using rules developed by Rubin.³ The way in which the plausible values for missing data are obtained is important, and in general, use of a misspecified imputation model results in invalid inferences. In general, it is desirable that the imputation model is compatible with the chosen substantive model. There exists a range of methods for performing MI covering different substantive model types; see Carpenter and Kenward⁴ for an overview. Two MI approaches have been described for imputation of missing data on covariates in Cox regression. White and Royston⁵ outlined an approximately compatible method, which can be implemented in standard software, and Bartlett et al⁶ described an alternative “substantive-model-compatible” approach, which does not require approximations.

Assessment of whether the covariate effect changes over time is the basis of a test of the proportional hazards assumption, which is an important aspect of model assessment in Cox regression. Furthermore, in time-to-event analyses, it is often of interest to study whether the association of certain covariates with the hazard changes over time. Ignoring time-varying effects (TVEs) and estimating an “average” hazard ratio can result in misleading conclusions.⁷ Cox¹ described an extended version of his model to incorporate TVEs of covariates, and there is a large literature on methods for estimating and testing for TVEs in Cox regression: chapter 6 in Therneau and Grambsch⁸ summarises some of the more popular methods. There is also a more recent literature on model building in Cox regression incorporating TVEs.^{9–13}

The existing imputation methods for handling missing data in Cox regression^{5,6} do not account for TVEs of covariates, which could result in invalid inferences. In this paper, we extend the methods of White and Royston⁵ and Bartlett et al⁶ to accommodate imputation of covariates modelled with TVEs in the Cox regression model. The methods are presented for a general form for a TVE. Specific details are given for TVEs modelled using restricted cubic splines, which are flexible and do not require a form for the TVE to be prespecified. We also present a model selection algorithm that incorporates imputation of missing data into a procedure for testing for proportional hazards and selecting a flexible functional form for TVEs. Throughout, we make the assumption that data are “missing at random” (MAR).^{3,14} Although the term “time-varying effect” is used, we note that a hazard ratio changing over time does not necessarily correspond to a covariate’s causal effect changing over time. It may instead occur when the association between a baseline covariate and the hazard becomes weaker (for example) over time or due to time-varying confounding or omitted covariates (see Collett¹⁵ and Ford and Norrie¹⁶ for details).

The paper is organised as follows. In Section 2, we describe extensions to the methods of White and Royston⁵ and Bartlett et al⁶ to accommodate TVEs, for a general functional form for the TVEs. Use of restricted cubic splines to model the TVEs is described in Section 3. In Section 4, we discuss testing the proportional hazards assumption and present a model selection algorithm. The proposed methods are investigated using simulation studies described in Section 5, in which several underlying functional forms for the TVEs are considered. The methods are illustrated using data from the Rotterdam Breast Cancer Study in Section 6, and we conclude with a discussion in Section 7. R code for implementing the methods is available at <https://github.com/ruthkeogh/MI-TVE>.

2 | MI FOR COX REGRESSION WITH TIME-VARYING EFFECTS (TVEs)

2.1 | Preliminaries

Let T denote the earlier of the event and censoring time and D be an indicator of whether an individual had the event ($D = 1$) or was censored ($D = 0$). For simplicity, we focus on a single covariate X_1 with missing data and a fully observed covariate, X_2 . Extensions to missingness in several variables are described in the Supplementary Materials (Section S2). Under the extended Cox model that allows TVEs of covariates,^{1,8} the hazard function can be written in the general form

$$h(t|X_1, X_2) = h_0(t) \exp \{ f_{X_1}(t; \beta_1)X_1 + f_{X_2}(t; \beta_2)X_2 \}, \quad (1)$$

where $h_0(t)$ is the baseline hazard and the potential TVEs for X_1 and X_2 are represented respectively by the functions $f_{X_1}(t; \beta_1)$ and $f_{X_2}(t; \beta_2)$. Under the standard Cox proportional hazards model, ie, with no TVEs, $f_{X_1}(t; \beta_1) = \beta_1$ and $f_{X_2}(t; \beta_2) = \beta_2$.

2.2 | MI overview

Using MI, the general procedure for obtaining estimates of the model parameters β_1 and β_2 is as follows (page 39 in the work of Carpenter and Kenward⁴). A model $p(X_1|T, D, X_2; \alpha)$ with noninformative prior on parameters α is specified for $p(X_1|T, D, X_2)$, the distribution of X_1 given T, D , and X_2 . Then, for $m = 1, \dots, M$,

1. The model $p(X_1|T, D, X_2; \alpha)$ is fitted by maximum likelihood to the data on individuals with observed X_1 to obtain an approximate posterior distribution for α . A value $\alpha^{(m)}$ is drawn from this distribution.
2. For each individual i with missing X_{1i} , a value $X_{1i}^{(m)}$ is drawn from $p(X_{1i}|T_i, D_i, X_{2i}; \alpha^{(m)})$, giving an “imputed” data set in which there are no missing values.
3. The substantive model, here, the Cox regression model, is fitted to this imputed data set to give estimates $(\hat{\beta}_1^{(m)}, \hat{\beta}_2^{(m)})$ of (β_1, β_2) and a corresponding estimate $\hat{\Sigma}^{(m)}$ of $\text{Var}(\hat{\beta}_1^{(m)}, \hat{\beta}_2^{(m)})$.

Estimates $(\hat{\beta}_1^{(m)}, \hat{\beta}_2^{(m)})$ ($m = 1, \dots, M$) and $\hat{\Sigma}^{(m)}$ are combined using “Rubin's rules” to give an overall estimate of (β_1, β_2) and corresponding variance-covariance matrix.³

The main difficulty that arises when the substantive model is a Cox regression is that $p(X_1|T, D, X_2)$ is a nonstandard distribution, which is only semiparametrized since $h_0(t)$ is nonparametric; therefore, we cannot easily draw values from the distribution $p(X_1|T, D, X_2)$ to obtain the imputations. Although in principle any model $p(X_1|T, D, X_2; \alpha)$ could be used, potentially serious (asymptotic) bias in the estimates of (β_1, β_2) and their variance-covariance matrix could arise if this model is misspecified. In particular, assuming the substantive model is correctly specified, if the imputation model is not compatible with the substantive model, under certain conditions, this implies that the imputation model is misspecified.⁶ Consequently, it is desirable that these two models be compatible (or approximately compatible), ie, that there exists a model for the joint distribution $p(X_1, T, D|X_2)$ that implies as submodels the model $p(X_1|T, D, X_2; \alpha)$ used for imputation and the Cox model used for analysis. Two different approaches to using a compatible or approximately compatible imputation model have been described by Bartlett et al⁶ and White and Royston,⁵ which we refer to respectively as MI-SMC and MI-Approx. In the next two subsections, we describe extensions of these imputation methods to accommodate TVEs in the Cox regression model.

2.3 | MI-TVE-Approx

For the standard Cox proportional hazards model assuming no TVEs ($f_{X_1}(t; \beta_1) = \beta_1$ and $f_{X_2}(t; \beta_2) = \beta_2$), White and Royston⁵ showed that an approximately compatible imputation model for X_1 is a logistic regression (for binary X_1) or linear regression (for continuous X_1) with linear predictor including main effects of $D, X_2, \hat{H}(T)$ and the interaction between X_2 and $\hat{H}(T)$, where $\hat{H}(T)$ is the Nelson-Aalen estimate of the cumulative hazard. Investigations have found, in the settings examined, that the interaction term adds little.^{5,17}

When the substantive model is the extended Cox model with TVEs in 1, we can show that an approximately compatible imputation model for X_1 is a logistic regression (for binary X_1) or linear regression (for continuous X_1) with linear predictor including main effects of $X_2, Df_{X_1}(T), \hat{H}(T), \hat{H}^{(1)}(T)$ and interactions of X_2 with $\hat{H}(T)$ and $\hat{H}^{(1)}(T)$, where $\hat{H}^{(1)}(T)$ is the Nelson-Aalen-type estimator $\hat{H}^{(1)}(T) = \sum_{t \leq T} \frac{td(t)}{n(t)}$ ($d(t)$ and $n(t)$ denote the number of events and number at risk at time t) and $f_{X_1}(T)$ denotes a vector of functions used for the TVE, eg if $f_{X_1}(t; \beta_1) = \beta_{10} + \beta_{11}t$ then $f_{X_1}(T) = (1, T)'$. The details of the derivation are given in the Supplementary Materials (Section S1). We refer to the resulting approach as MI-TVE-Approx. In the simulations, we will investigate whether it is important to include the higher-order cumulative hazard term $\hat{H}^{(1)}(T)$ and/or the interaction terms $X_2 \times \hat{H}(T)$ and $X_2 \times \hat{H}^{(1)}(T)$. The imputation procedure is as follows.

1. Fit the imputation model to the subset of individuals with complete data. For continuous X_1 , this is

$$X = \alpha_0 + \alpha_1 X_2 + \alpha_2' Df_{X_1}(T) + \alpha_3 \hat{H}(T) + \alpha_4 \hat{H}^{(1)}(T) + \alpha_5 X_2 \hat{H}(T) + \alpha_6 X_2 \hat{H}^{(1)}(T) + \epsilon,$$

and for binary X_1 ,

$$\text{logit Pr}(X_1 = 1|T, D, X_2) = \alpha_0 + \alpha_1 X_2 + \alpha_2' Df_{X_1}(T) + \alpha_3 \hat{H}(T) + \alpha_4 \hat{H}^{(1)}(T) + \alpha_5 X_2 \hat{H}(T) + \alpha_6 X_2 \hat{H}^{(1)}(T).$$

2. Take M random draws values of the parameters from their approximate posterior distribution (we refer to the work of White et al¹⁸ for details), denoted $\alpha_j^{(m)}$ ($j = 0, 1, \dots, 6$) (binary and continuous X_1), and additionally $\sigma_\epsilon^{2(m)}$ for continuous X_1 , where σ_ϵ^2 denotes the variance of the residuals ϵ .

3. The imputed value of X_{1i} in the m th imputed data set is given (for continuous X_1) by

$$X_{1i}^{(m)} = \alpha_0^{(m)} + \alpha_1^{(m)} X_2 + \alpha_2^{(m)'} Df_{X_1}(T) + \alpha_3^{(m)} \hat{H}(T) + \alpha_4^{(m)} \hat{H}^{(1)}(T) + \alpha_5^{(m)} X_2 \hat{H}(T) + \alpha_6^{(m)} X_2 \hat{H}^{(1)}(T) + \varepsilon^{(m)},$$

where $\varepsilon^{(m)}$ is a random draw from a normal distribution with mean 0 and variance $\sigma_\varepsilon^{2(m)}$. For binary X_1 , the imputed value is a draw from a Bernoulli distribution with

$$\text{logit Pr}(X_1 = 1|T, D, X_2) = \alpha_0^{(m)} + \alpha_1^{(m)} X_2 + \alpha_2^{(m)'} Df_{X_1}(T) + \alpha_3^{(m)} \hat{H}(T) + \alpha_4^{(m)} \hat{H}^{(1)}(T) + \alpha_5^{(m)} X_2 \hat{H}(T) + \alpha_6^{(m)} X_2 \hat{H}^{(1)}(T).$$

2.4 | MI-TVE-SMC

In the context of the standard Cox proportional hazards model without TVEs, MI-Approx has been found to work well in a range of circumstances.^{4,5} However, the approximation can perform badly in some “extreme” situations, including when there are large effect sizes and when the event rate is high.⁵ Bartlett et al⁶ described an approach, referred to here as MI-SMC, which ensures that the imputation model is compatible with the user’s chosen substantive model; here, a Cox regression (“substantive-model-compatible” – SMC). However, they did not accommodate TVEs.

To extend the method of Bartlett et al⁶ to modelling of TVEs, we adapt the approach first proposed in Bartlett’s PhD thesis¹⁹ for handling missing data in time-dependent covariates, noting the close connection between time-dependent covariates and TVEs. We refer to the resulting method as MI-TVE-SMC. The MI-TVE-SMC imputation procedure is as follows.

First, a model $p(X_1|X_2; \gamma)$ is specified. For binary X_1 , this may be a logistic regression model, and for continuous X_1 , a linear regression model. The steps used to obtain the m th imputed data set are then as follows.

1. Fill in the missing variables with arbitrary starting values to create a complete data set.
2. Fit the Cox regression model of interest, including TVEs, to the current complete data set to obtain estimates $(\hat{\beta}_1, \hat{\beta}_2)$ and their estimated variance-covariance matrix $\hat{\Sigma}$. Draw values $\beta_1^{(m)}, \beta_2^{(m)}$ from a joint normal distribution with mean $(\hat{\beta}_1, \hat{\beta}_2)$ and variance-covariance matrix $\hat{\Sigma}$.
3. Calculate Breslow’s estimate,²⁰ denoted $H_0^{(m)}(t)$, of the baseline cumulative hazard $H_0(t)$ using parameter values $\beta_1^{(m)}, \beta_2^{(m)}$ and current imputations of X_1 .
4. Estimate parameters γ and their variance-covariance by fitting the assumed regression model for X_1 on X_2 to the current complete data set. Draw a value γ^* from the approximate joint posterior distribution of γ .¹⁸
5. For each individual for whom X_1 is missing, (i) draw a value X_1^* from the distribution $p(X_1|X_2; \gamma^*)$ and (ii) draw a value U from a uniform distribution on $[0, 1]$. Accept the value X_1^* if

$$\begin{cases} U \leq \exp \left[-\sum_{j:t_j \leq T} \Delta H_0^{(m)}(t_j) \exp \left\{ f_{X_1} \left(t_j; \beta_1^{(m)} \right) X_1^* + f_{X_2} \left(t_j; \beta_2^{(m)} \right) X_2 \right\} \right] & \text{if } D = 0 \\ U \leq \Delta H_0^{(m)}(T) \exp \left\{ 1 + f_{X_1} \left(T; \beta_1^{(m)} \right) X_1^* + f_{X_2} \left(T; \beta_2^{(m)} \right) X_2 \right. \\ \left. - \sum_{j:t_j \leq T} \Delta H_0^{(m)}(t_j) e^{f_{X_1} \left(t_j; \beta_1^{(m)} \right) X_1^* + f_{X_2} \left(t_j; \beta_2^{(m)} \right) X_2} \right\} & \text{if } D = 1, \end{cases}$$

where $\Delta H_0^{(m)}(t)$ denotes the increment in $H_0^{(m)}(t)$ at time t and t_1, \dots, t_k denote the unique event times. Repeat (i) and (ii) until a value X_1^* is accepted.

6. Return to steps 2–5 until the imputed X_1 values have converged to a stationary distribution. These are then the imputed values in the m th imputed data set.

The difference between the MI-SMC approach, which does not accommodate TVEs, and the MI-TVE-SMC approach is in the terms used for the rejection in step 5 and the fact that a Cox model with TVEs is fitted in Step 2.

We have outlined the MI-TVE-Approx and MI-TVE-SMC approaches for the simple setting of missing data in a single covariate X_1 with a TVE. Both methods extend to handle missingness in several covariates using the fully conditional specification approach (also referred to as multiple imputation by chained equations), in which an imputation model is specified for each partially missing covariate conditional on all the other covariates and an iterative approach is used to fit the imputation models.^{18,21} Details are provided in the Supplementary Materials (Section S2).

3 | FUNCTIONAL FORM OF TIME-VARYING EFFECTS (TVEs)

In the preceding section, the MI methods were described for a general functional form for the TVEs, $f_X(t; \beta)$. Approaches to modelling TVEs include use of prespecified functional forms²² (eg, $f_X(t; \beta) = \beta_0 + \beta_1 t$), step-functions,²²⁻²⁴ fractional polynomials,¹² and splines.^{10,11,25-30} In this paper, we focus on using a restricted cubic spline form^{25,26} for the TVE function because this allows a flexible form for the TVE with relatively few parameters. Under a restricted cubic spline with L knots at u_1, \dots, u_L , the TVE function for a covariate X is

$$f_X(t; \beta) = \beta_0 + \beta_1 t + \sum_{i=1}^{L-2} \theta_i \left\{ (t - u_i)_+^3 - \left(\frac{(t - u_{L-1})_+^3 (u_L - u_i)}{(u_L - u_{L-1})} \right) + \left(\frac{(t - u_L)_+^3 (u_{L-1} - u_i)}{(u_L - u_{L-1})} \right) \right\}, \quad (2)$$

where $(t - u_i)_+$ takes value $(t - u_i)$ if $(t - u_i) > 0$ and 0 otherwise, and β includes β_0, β_1 and $\theta_i (i = 1, \dots, L-2)$. The number of knots used and the position of the knots have to be decided by the user, and there is no formal theoretical basis for the decision. Hess²⁵ noted empirical evidence that three to five knots are usually adequate and the fit is not greatly altered by altering the knot positions. Stone³¹ also recommended using five knots in restricted cubic splines. Hess²⁵ suggested placing knots at quantiles of the observed follow-up times (including both event times and censoring times), with the outer knots near the extremes, and the internal knots placed approximately uniformly over the quantiles of the distribution of the follow-up times. Similar recommendations were given by Durrleman and Simon³² in the context of restricted cubic splines for functional forms of covariates in survival analyses. In the simulations, we consider using restricted cubic splines with five knots placed at percentiles (5th, 25th, 50th, 75th, 95th) of the event time distribution (ie, excluding censoring times).

In the case of a restricted cubic spline with $L = 5$ knots, the MI-TVE-Approx imputation model for X_1 should include X_2, D , the interaction between D and T , interactions of D with $\left\{ (T - u_i)_+^3 - \frac{(T - u_{L-1})_+^3 (u_L - u_i)}{(u_L - u_{L-1})} + \frac{(T - u_L)_+^3 (u_{L-1} - u_i)}{(u_L - u_{L-1})} \right\}$ for $i = 1, 2, 3$, $\hat{H}(T)$, $\hat{H}^{(1)}(T)$, and interactions of X_2 with $\hat{H}(T)$ and $\hat{H}^{(1)}(T)$.

4 | TESTING THE PROPORTIONAL HAZARDS ASSUMPTION AND MODEL SELECTION

In most contexts, when using Cox regression modelling, it is important to assess whether covariates have TVEs, that is, to perform tests of the proportional hazards assumption. TVEs can then be included for covariates for which the proportional hazards assumption appears not to hold. Tests of proportional hazards based on TVEs modelled using splines have been previously described by Abrahamowicz et al.²⁷ With fully observed data on covariates, the proportional hazards assumption can be tested using a likelihood ratio test, comparing a model including TVEs to the model without TVEs. A joint Wald test of the TVE parameters is asymptotically equivalent: assuming a restricted cubic spline for the TVE for X (Equation 2), this is a joint test of $\beta_1 = \theta_1 = \dots = \theta_{L-2} = 0$. Wood et al³³ described the use of Wald tests for model selection using multiply imputed data, and this was further evaluated by Morris et al³⁴ in the context of covariate transformations based on fractional polynomials. We suggest this approach for tests of TVEs. The joint test of the parameters of interest (null hypothesis $\beta_1 = \theta_1 = \dots = \theta_{L-2} = 0$) uses the pooled parameter estimates and the corresponding variance covariance matrix obtained from Rubin's rules. We use the Wald-type test described by Li et al³⁵ and Meng and Rubin³⁶ as this was found to perform slightly better in terms of giving correct type I errors in subsequent simulation studies (Section 5).

For the purposes of testing the proportional hazards assumption as part of a model assessment and selection procedure, we recommend allowing TVEs for all variables at the imputation stage of the analysis; the importance of doing so for valid tests of the proportional hazards assumption is investigated in the simulation studies.

Several authors have proposed algorithms for model selection involving both TVEs and transformation of covariates,¹⁰⁻¹² though all assume fully observed data. We propose an algorithm (the MI-MTVE algorithm), which provides a model selection procedure for identifying TVEs using multiply imputed data. The MI-MTVE algorithm is an adaptation of the "multivariable fractional polynomial-time" (MFPT) algorithm of Sauerbrei et al,¹² which uses fractional polynomial transformations of covariates and fractional polynomial forms for TVEs. Our adaptation employs restricted cubic splines rather than fractional polynomials for TVEs, and is similar to a procedure advocated by Wynant and Abrahamowicz.¹¹ Forward selection is used to accommodate investigation of TVEs in multiple covariates and selection of functional forms for TVEs using restricted cubic splines with up to five knots.

MI-MTVE algorithm

- Step 1** Perform MI-TVE-Approx or MI-TVE-SMC, assuming a restricted cubic spline with five knots for the TVE for each covariate to obtain M imputed data sets.
- Step 2** In each imputed data set, fit the Cox regression model with no TVEs of any covariate (denoted model \mathcal{M}_0). Denote the set of covariates by C . For each $c (c \in C)$ and in each imputed data set, fit four Cox regression models including TVEs for covariate c of increasing complexity (indexed by j): using the linear form ($f_X(t; \beta) = \beta_0 + \beta_1 t$) and restricted cubic splines with three, four, and five knots. The remaining covariates in the Cox regression are not modelled using TVEs.
- Step 3** For each covariate c , perform a test for TVEs in each model j using joint Wald tests for multiply imputed data, as described above.^{35,36} Select the combination of covariate (c) and TVE model (j) that returns the smallest p-value in the test for TVEs. If no combination of c and j gives a p-value less than a chosen level α , stop; the working model without TVEs \mathcal{M}_0 is adequate. Otherwise, update the working model \mathcal{M}_0 to include TVEs for the covariate c using the TVE model j that returned the smallest p-value. Call this new working model \mathcal{M}_1 .
- Step 4** Repeat steps 2–3 with updated working models until there are no remaining covariates c not in the current working model that have a significant TVE (at level α) under any TVE model j . Stop; this working model is the final selected model.

The estimates of the parameters of the final selected model and corresponding estimated variance-covariance matrix are those obtained by applying Rubin's rules to the results from fitting the final model to each imputed data set. The MI is performed only in step 1 and is based on a TVE for each covariate of the most complex form that we consider in this paper (a restricted cubic spline with five knots). This means that a restriction of the imputation model should be compatible with the model selected by the algorithm (termed "semi-compatibility").^{6,34,37} The imputation model may include some redundant parameters, but this will not impact on the validity of MI inference. Collins et al³⁸ showed that the cost in terms of efficiency of including many predictors in the imputation model (ie, with potentially many redundant parameters) is relatively small and that this is outweighed by the benefits in terms of potential to avoid bias and actually gain efficiency.

5 | SIMULATION STUDY

We now present a simulation study which was designed to evaluate the MI methods across a variety of data-generating scenarios.

5.1 | Data-generating mechanisms

Data were generated for a cohort of 2000 individuals. Two covariates, X_1 and X_2 , are considered. Event times T_E were generated according to the exponential hazard model

$$h(t|X_1, X_2) = \lambda_E \exp \{ f_{X_1}(t; \beta_1)X_1 + f_{X_2}(t; \beta_2)X_2 \}. \quad (3)$$

We consider five forms for the TVEs. These are listed in the table at the top of Figure 1. In scenario 1, neither covariate has a TVE. In scenarios 2–5, X_1 has a TVE but X_2 does not. Figure 1 depicts the form of the TVEs. These forms include examples previously used by Buchholz³⁹ and Buchholz and Sauerbrei.⁹

Random drop out times, T_C , were generated according to an exponential distribution with rate λ_C , and administrative censoring was imposed after 10 years of follow-up. The observed time for each individual was calculated as $T = \min(T_E, T_C, 10)$. Values for λ_E and λ_C were chosen such that 10% of individuals have the event of interest and 50% are censored because of random drop out, with the remainder being administratively censored.

Both binary and continuous pairs of covariates are considered. Binary X_1 was generated from a binomial distribution such that $P(X_1 = 1) = 0.2$, and binary X_2 was generated using $\text{logit}\{P(X_2 = 1|X_1)\} = X_1$. Continuous X_1 and X_2 were generated from a bivariate normal distribution with means 0, variances 1, and correlation 0.5.

Nonmonotone missing data were generated in X_1 and X_2 according to a MAR mechanism in which the probability of missingness in X_1 depends on observed values of X_2 and vice versa (see Supplementary Materials Section S4). In this, X_1 is missing for 30% of individuals and X_2 for 30% of individuals, resulting in approximately 50% of individuals missing at least one of the measurements.

There are 10 main simulation scenarios: five different scenarios for TVEs of X_1 and binary and continuous X_1 and X_2 . Five hundred simulated data sets were generated under each scenario (justified in the Supplementary Materials

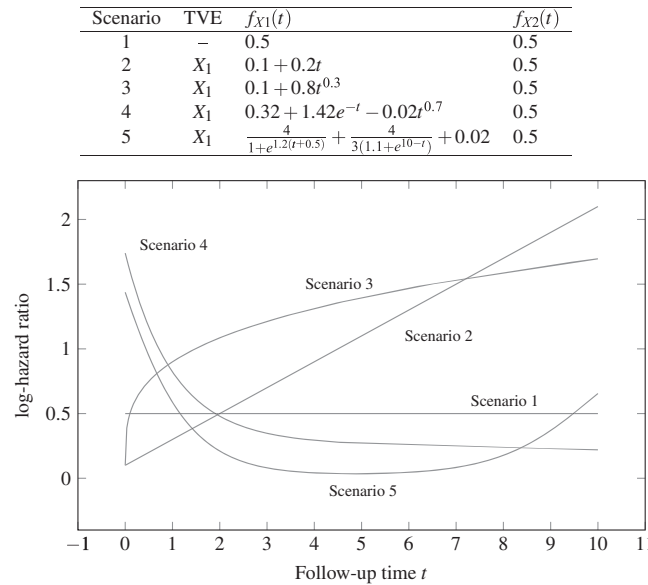


FIGURE 1 Time-varying effect (TVE) functions used in simulation studies

Section S5). In Section 5.5, we present results from additional sensitivity scenarios with a higher event rate, lower level of missingness, and a MAR mechanism in which the missingness in X_1 and X_2 additionally depends on the outcome D .

5.2 | Methods compared

The methods we investigate are the proposed MI-TVE-Approx and MI-TVE-SMC approaches, and for comparison, the corresponding approaches that do not incorporate TVEs (MI-Approx and MI-SMC). We also performed a complete-data analysis (before missing data is introduced) and a complete-case analysis, which uses only the subset with no missing data. In MI-Approx, we omitted the interaction terms between covariates and $\hat{H}(T)$ because their inclusion was not found to result in material differences in the results. For the same reason, in MI-TVE-Approx, we omitted the interaction terms and terms involving $\hat{H}^{(1)}(T)$. The MI-TVE-Approx imputation model recommended for X_1 therefore includes X_2 , D , the interaction between D and T , interactions of D with $\left\{ (T - u_i)_+^3 - \frac{(T - u_{L-1})_+^3 (u_L - u_i)}{(u_L - u_{L-1})} + \frac{(T - u_L)_+^3 (u_{L-1} - u_i)}{(u_L - u_{L-1})} \right\}$ for $i = 1, 2, 3$, and $\hat{H}(T)$. The recommended imputation model for X_2 is the same but with X_2 replaced by X_1 .

In all Cox regression analyses, TVEs for X_1 and X_2 are modelled using a restricted cubic spline with five knots placed at percentiles (5th, 25th, 50th, 75th, 95th) of the distribution of the observed event times. This includes a TVE model for X_2 ($f_{X_2}(t; \beta_2)$) even though, in the data-generating process, there is no TVE of X_2 . In MI-TVE-Approx and MI-TVE-SMC, the TVE was incorporated on the basis of the same functional form.

In the MI analyses, we used 10 imputed data sets. For the analysis of studies in practice, we recommend the rule of thumb suggested by White et al¹⁸ to set the number of imputations to be approximately the same as the percentage of missing data, with a larger number chosen if numerical reproducibility of estimates is desired.

5.3 | Performance measures

The performance of methods was assessed in a number of ways as described below. Each assessment was performed for both X_1 and X_2 .

- Curvewise estimate of the TVE, presented visually over the follow-up time and averaged over simulation runs.
- Bias in the estimated curve at one, five, and nine years and corresponding 95% Monte Carlo confidence intervals. The bias from the MI methods and the complete-case analysis was calculated relative to the complete-data results, ie, as the difference between the MI or complete-case estimates and the mean of the complete-data estimates. This was done because the true data-generating mechanism is not a restricted cubic spline and therefore we do not necessarily expect to get completely unbiased estimates from the complete-data analysis.

- Coverage of confidence intervals, estimated at one, five, and nine years, defined as the proportion of simulated data sets for which the true curve lies within the 95% confidence intervals at time $t=1,5,9$.
- Curvewise root mean squared error (RMSE), presented visually over the follow-up time and averaged over simulation runs.
- Rejection fractions for the test of the proportional hazards assumption. For scenario 1, in which there are no TVEs, this corresponds to a type I error rate for the TVEs of both X_1 and X_2 . For all other scenarios, this corresponds to power for the TVE of X_1 and type I error rate for the TVE of X_2 . The proportional hazards assumption is assessed using a joint Wald test of the TVE parameters (see Section 4).

All simulations and analyses were performed using R. The substantive model was fitted using `coxph` in the `survival` package. MI-Approx and MI-TVE-Approx were implemented using `mice`⁴⁰ and MI-SMC using `smcfcs` (<https://github.com/jwb133/smcfcs>). We extended the `smcfcs` code to implement MI-TVE-SMC in R, and this is available at <https://github.com/ruthkeogh/MI-TVE>, together with simulation examples. The code allows the user to specify the TVE function.

5.4 | Simulation results

5.4.1 | Bias, coverage, and RMSE

Figures 2 and 3 show the curvewise TVE estimates for covariate X_1 in the binary and continuous covariates settings. Supplementary Figures S1 and S2 show the bias in the estimated TVE curves at three time points (1, 5, 9), corresponding to the difference between the mean curves shown in Figures 2 and 3 and the true curve or that obtained from the complete

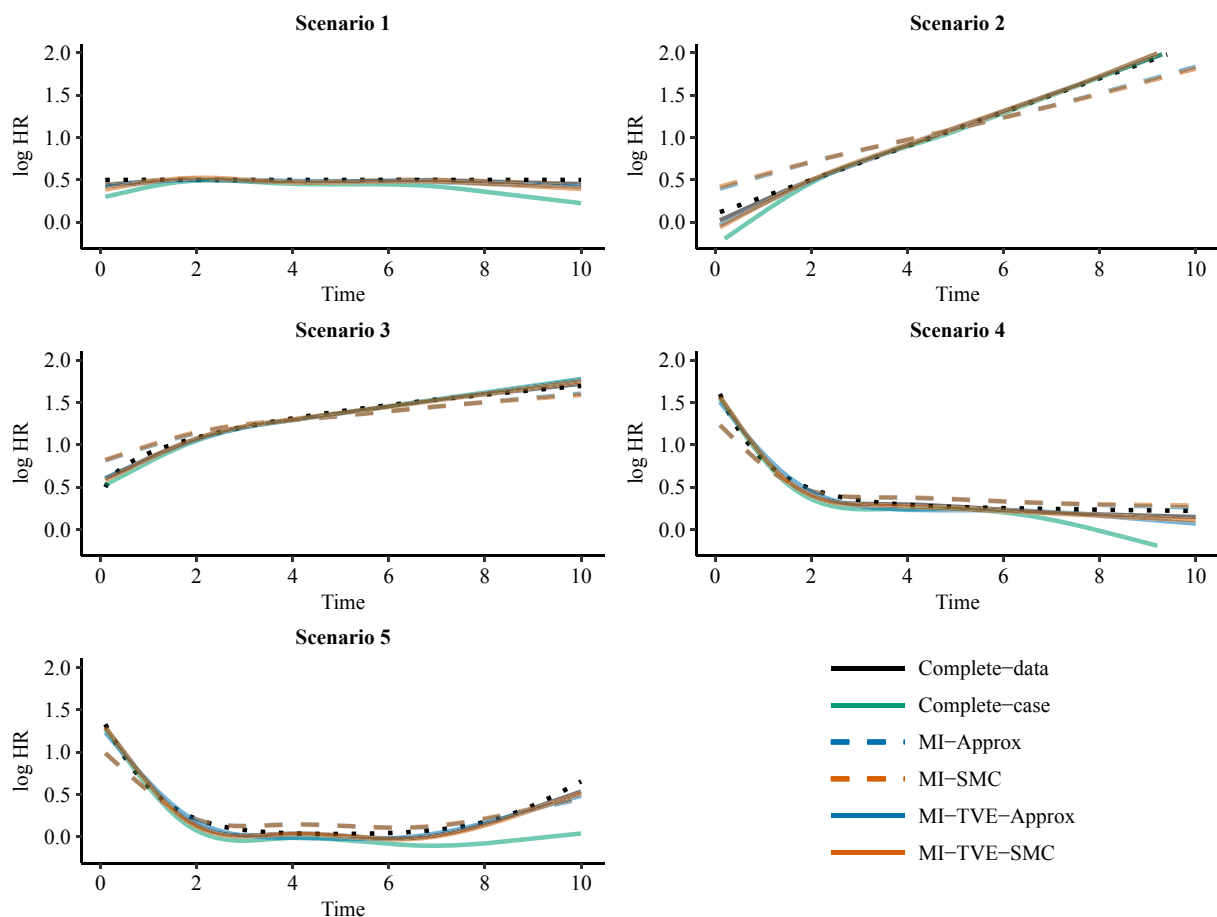


FIGURE 2 Curve-wise estimates of TVEs for covariate X_1 in the setting with binary covariates X_1 and X_2 . The dotted black line indicates the true curve. MI, multiple imputation; SMC, substantive-model-compatible; TVE, time-varying effect [Colour figure can be viewed at wileyonlinelibrary.com]

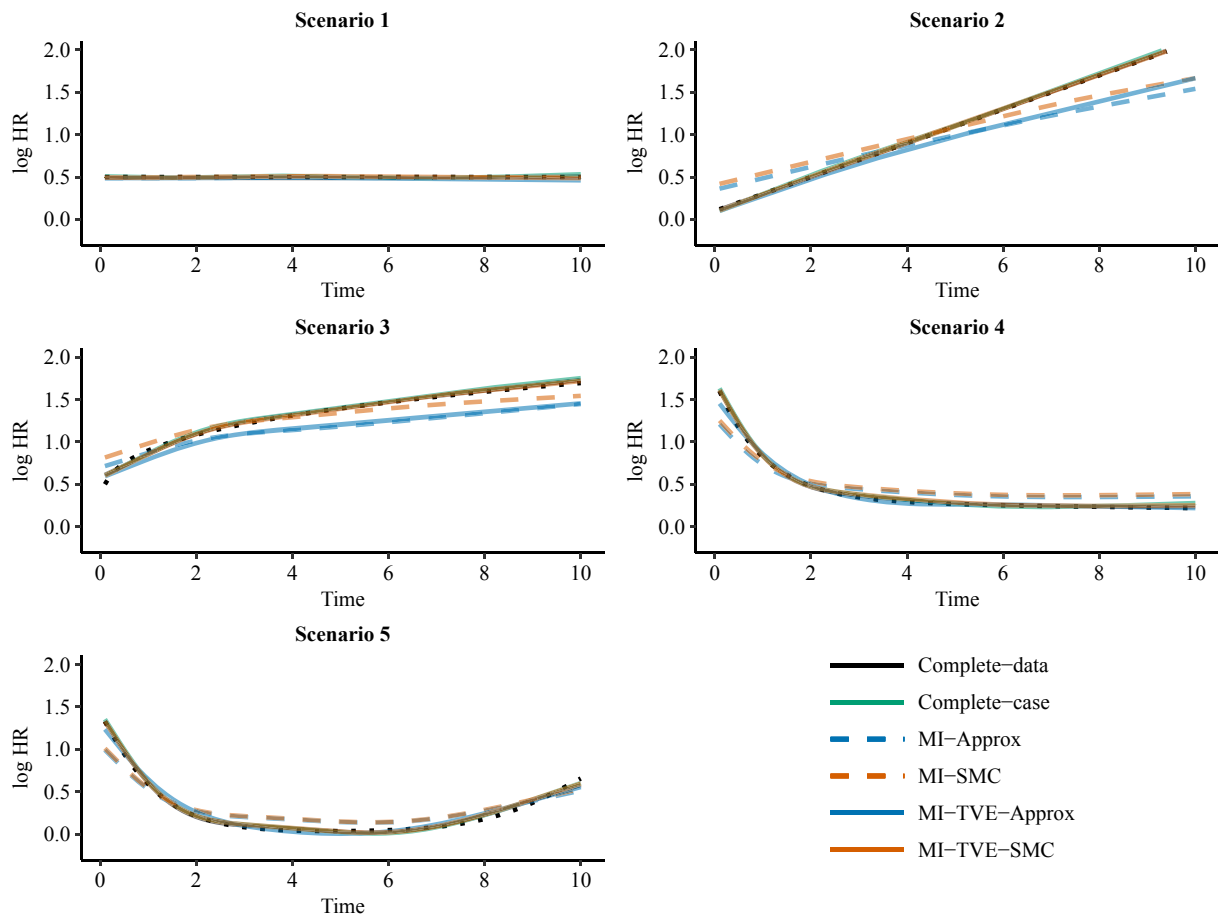


FIGURE 3 Curve-wise estimates of TVEs for covariate X_1 in the setting with continuous covariates X_1 and X_2 . The dotted black line indicates the true curve. MI, multiple imputation; SMC, substantive-model-compatible; TVE, time-varying effect [Colour figure can be viewed at wileyonlinelibrary.com]

data analysis. Similar plots for X_2 , which always has a time-constant effect, are shown in the Supplementary Materials (Figures S3 and S4).

The complete-data and complete-case analyses give approximately unbiased TVE estimates except for some bias in the complete-case analysis in the extremes of some curves. As noted earlier, the complete-data analysis could give estimates with some slight bias because the data were not generated under the restricted cubic spline model, which is used in the analysis. Note that we expect the complete-case analysis to give an approximately unbiased result because the missingness does not depend on the outcome. The results from MI-Approx and MI-SMC show that failing to account for the TVE in the imputation gives a biased estimate of the TVE curve. The bias is such that the TVE curve appears more flat and the HR is overestimated at some times and underestimated at others. For example, in Scenario 2, where the true TVE is linear, the slopes of the estimated TVE from MI-Approx and MI-SMC are less steep than the true slope. The MI methods that accommodate TVEs, MI-TVE-Approx and MI-TVE-SMC, perform similarly for binary X_1 and give TVE estimates similar to that from the complete-data analysis. However, for continuous X_1 , only MI-TVE-SMC performs well in general, with MI-TVE-Approx giving clearly biased estimates in scenarios 2 and 3 at times where the TVE is quite large. MI-TVE-Approx requires additional approximations for continuous covariates, and the approximation does not perform well in these scenarios. Poor performance of MI-Approx in scenarios with continuous covariates and large effect sizes has been found previously in the setting without TVEs.⁵

Tables of coverages of the estimated TVE curves at three time points are given in Supplementary Tables S1 and S2. The coverage levels tend to be higher than the nominal 95% level at the later timepoints, with many reaching 100%. This is not an issue specific to MI since we see the same behaviour for the analysis of complete data; it is at least reassuring that coverage results for the MI-TVE-SMC method follow a very similar pattern to those from the complete-data analysis. Coverage not at the nominal level has been previously observed for spline-based models.⁴¹

TABLE 1 Percentage of simulations in which the null hypotheses of proportional hazards for X_1 and X_2 were rejected using joint Wald tests^{35,36} for binary and continuous covariates.For a given rejection percentage π , the Monte Carlo SE is $\sqrt{\frac{\pi(100-\pi)}{500} \times \frac{1}{100}}$

	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5	
	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
<i>Binary X_1, X_2</i>										
Complete-data	3	3	89	3	33	3	56	6	45	4
Complete-case	2	3	42	3	14	2	17	2	14	3
MI-Approx	0	0	21	0	3	0	4	0	2	0
MI-SMC	0	0	17	0	2	0	3	0	1	0
MI-TVE-Approx	2	3	67	3	16	2	27	2	21	2
MI-TVE-SMC	3	4	68	6	24	6	34	5	27	5
<i>Continuous X_1, X_2</i>										
Complete-data	7	5	100	3	78	5	99	3	96	4
Complete-case	5	6	94	5	42	4	78	4	60	5
MI-Approx	0	0	70	0	7	0	40	0	28	0
MI-SMC	0	0	82	0	9	0	48	0	30	0
MI-TVE-Approx	4	5	91	2	23	2	83	3	70	3
MI-TVE-SMC	5	7	99	4	53	5	88	3	75	5

Abbreviations: MI, multiple imputation; SMC, substantive-model-compatible; TVE, time-varying effect.

Supplementary Figures S5 and S6 show curvewise RMSE. As we expect, the complete-case analysis gives the largest RMSE. The RMSEs from MI-TVE-Approx and MI-TVE-SMC are slightly larger than those from the complete-data analysis and from the MI analyses that ignore TVEs. However, these plots show that there is not a substantial bias-variance trade-off in the MI analyses that accommodate TVEs. Supplementary Figures S7 and S8 further illustrate this point by showing estimated curves from 100 example simulated data sets using the complete-data analysis and MI-TVE-SMC: the curves obtained using MI-TVE-SMC have a greater spread than those from the complete-data analysis, as we would expect, but not greatly so.

5.4.2 | Tests of the proportional hazards assumption

Table 1 shows the percentage of simulations in which the proportional hazards assumption (based on joint Wald tests) was rejected at the 5% level for X_1 and X_2 in the binary and continuous covariates settings. In scenario 1, where neither X_1 nor X_2 has a TVE, the percentage of simulations in which the null hypothesis of proportional hazards was rejected was close to 5% in the complete-data and complete-case analyses, and also in MI-TVE-Approx and MI-TVE-SMC, indicating approximately correct type I errors. However, the results of zero from MI-Approx and MI-SMC show that failing to account for the possibility of TVEs in the imputation model will result in incorrect type I errors. Type I error results are similar for X_2 , which has a time-constant effect across scenarios. In scenarios 2–5 with TVEs for X_1 , the power to reject the null hypothesis of proportional hazards varied under the complete-data analysis, from 100% (continuous covariates, scenario 2) to 33% (binary covariates, scenario 3). Power was generally lower in the setting with binary covariates. Power was reduced under the complete-case analysis; for example, in scenario 4 with continuous covariates, the power from the complete-case analysis was 78% compared with 99% in the complete-data analysis, and in scenario 4 with binary covariates, the power from the complete-case analysis was 17% compared with 56% in the complete-data analysis. For binary covariates, the power under MI-TVE-Approx and MI-TVE-SMC was increased relative to the complete-case analysis and was highest for MI-TVE SMC. Power using MI-TVE-SMC was also high in the setting with continuous covariates but lower for MI-TVE-Approx and, in scenarios 2 and 3, lower than that from the complete-case analysis; this is partly a consequence of the bias observed using MI-TVE-Approx for continuous covariates. The results show that if the TVEs are ignored in the imputation (MI-Approx and MI-SMC), there is a large loss of power to reject the null hypothesis of proportional hazards across all scenarios, and power from these methods was considerably lower than that from the complete-case analysis.

5.5 | Additional simulation investigations

We investigated the performance of the methods in three additional situations.

A. Missingness depends on the outcome (see Supplementary Materials Section S4). Missingness depending on the

outcome is plausible if there is an underlying latent feature that is associated with the subsequent outcome and with missingness.

- B. Fifty percent of individuals have the event. This may not be a common situation but is relevant for certain clinical studies.
- C. A lower percentage of individuals with missing covariate data. The percentage of individuals missing X_1 and missing X_2 was reduced to 10% (see Supplementary Materials Section S4), which results in approximately 20% of individuals missing at least one of the measurements.

Other aspects of the simulations were as described above. Results are presented in Figure 4. For additional simulations (A) and (B), results are presented for scenario 4 with binary covariates, representing a situation in which the association between X_1 and the hazard becomes weaker over time. For additional simulation (C), we focused on scenario 2 with continuous covariates, for which we found biased estimates using MI-TVE-Approx in the earlier simulation results.

When the missingness depends on the outcome the complete-case analysis gives biased estimates. The results show that the proposed MI methods continue to perform well, while ignoring TVEs in the imputation still results in bias, as we would expect on the basis of our earlier results. The proposed methods also continue to perform well in a situation in

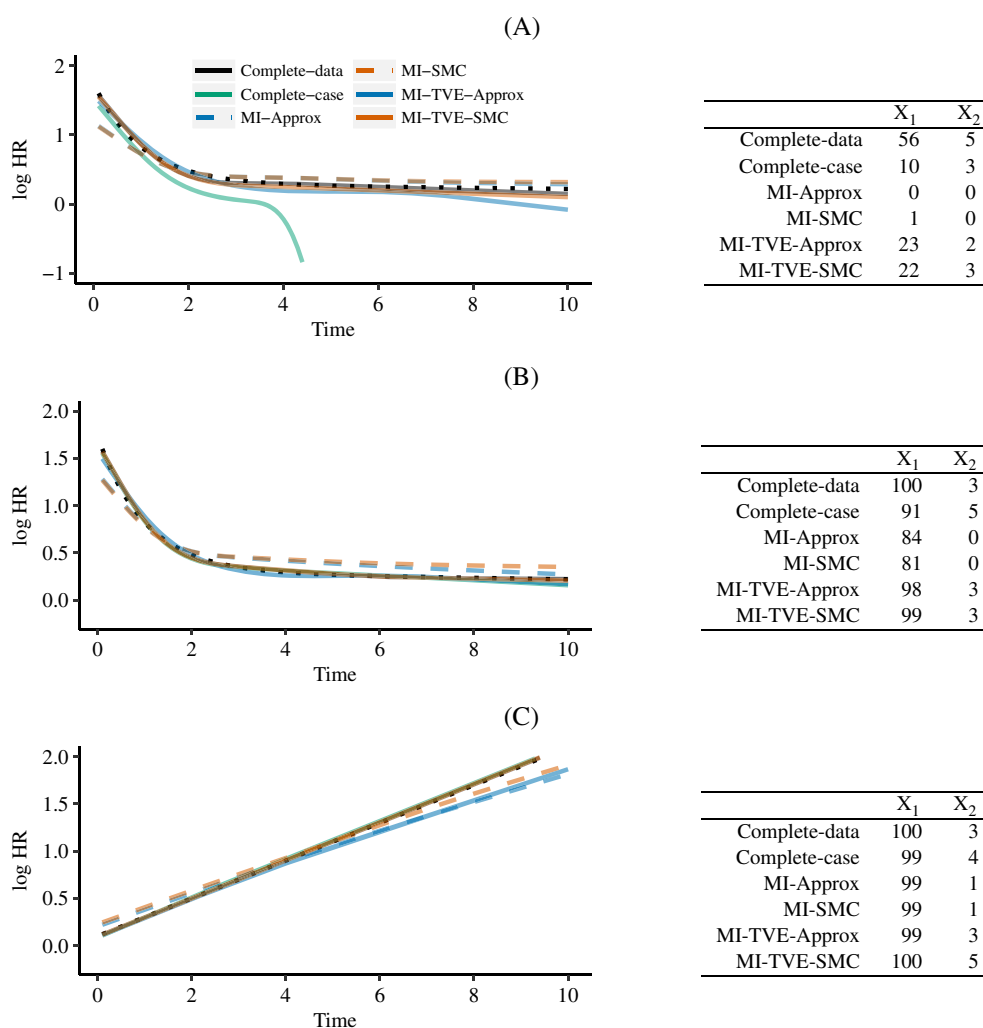


FIGURE 4 Results from additional simulations. The plots show the curvewise estimates of TVEs for covariate X_1 . The dotted black line indicates the true curve. The tables show the percentage of simulations in which the null hypotheses of proportional hazards for X_1 and X_2 were rejected using joint Wald tests. A, Scenario 4, binary X_1 and X_2 : The probability of missingness in X_1 and X_2 depends on D; B, Scenario 4, binary X_1 and X_2 : 50% of individuals had the event; C, Scenario 2, continuous X_1 and X_2 : the proportion of individuals missing X_1 or X_2 was reduced to 20%. MI, multiple imputation; SMC, substantive-model-compatible; TVE, time-varying effect [Colour figure can be viewed at wileyonlinelibrary.com]

which 50% of individuals have the event. When the proportion of individuals with missing data is reduced, the bias from MI-TVE-Approx in scenario 2 with continuous covariates is reduced but still evident when the time-varying effect is large.

6 | ILLUSTRATION: ROTTERDAM BREAST CANCER STUDY

We illustrate the methods using data on 2982 individuals with primary breast cancer from the Rotterdam tumour bank. The data set is freely available (we used the data set provided at <https://www.imbi.uni-freiburg.de/Royston-Sauerbrei-book/index.html#datasets>). This data set was chosen for two principal reasons: (i) it has been used by other authors to illustrate methods for handling TVEs^{12,42}; (ii) the data set being freely available, readers can use our code provided at <https://github.com/ruthkeogh/MI-TVE> to verify the analysis and change it for their own purposes.

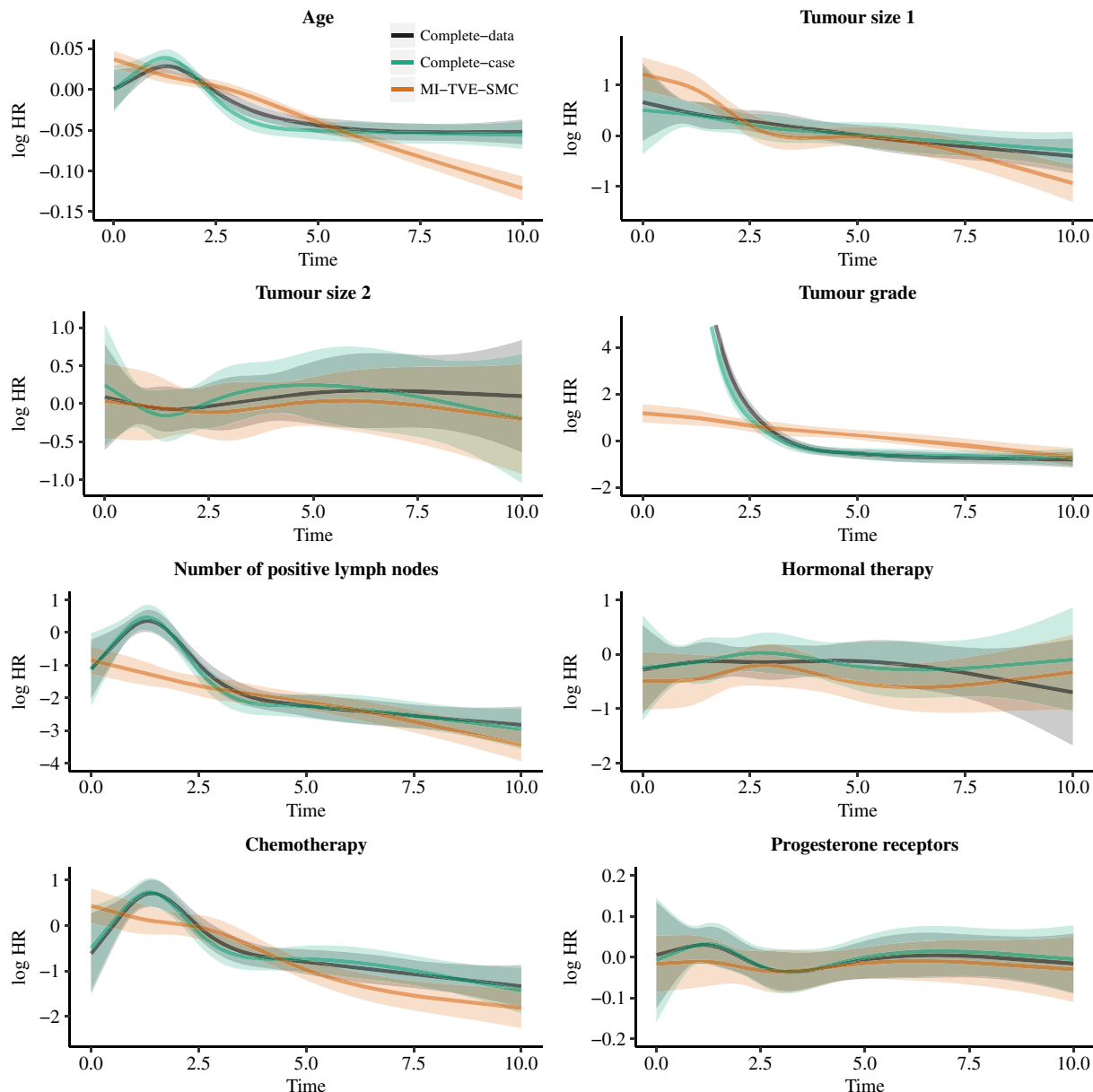


FIGURE 5 Results from the Rotterdam Breast Cancer Study. Plots showing estimated log hazard ratios as a function of time from a complete-data analysis, complete-case analysis, and MI-TVE-SMC analysis. The time-varying effects for all covariates were modelled using a restricted cubic spline with 5 knots. Shaded errors show 95% confidence intervals. Results are shown up to time 10. MI, multiple imputation; SMC, substantive-model-compatible; TVE, time-varying effect [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Results from the Rotterdam Breast Cancer Study. (i) P-value from a joint Wald test of the null hypothesis of no TVE for that covariate, based on a model in which TVEs were modelled using restricted cubic splines with 5 knots. (ii) Indication of whether TVEs for that covariate were included (✓) or excluded (✗) using the MI-MTVE algorithm, and if included, the functional form of the TVE. * 'lin' denotes a TVE of linear form $f_X(t; \beta) = \beta_0 + \beta_1 t$. "k3", "k4", and "k5" denote restricted cubic spline forms for the TVE with 3, 4, and 5 knots

Covariate	Complete-data		Complete-case		MI-Approx		MI-SMC		MI-TVE-Approx		MI-TVE-SMC	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
Age	< 0.001	✓(k5)	< 0.001	✓(k5)	< 0.001	✓(k5)	< 0.001	✓(k5)	< 0.001	✓(k5)	< 0.001	✓(k5)
Tumour size 1	< 0.001	✓(lin)	0.025	✓(lin)	< 0.001	✓(k5)	< 0.001	✓(k3)	< 0.001	✓(k5)	< 0.001	✓(k5)
Tumour size 2	0.941	✗	0.678	✗	0.077	✓(k3)	0.089	✗	0.850	✓(k3)	0.086	✓(k3)
Tumour grade	< 0.001	✓(k5)	< 0.001	✓(k5)	< 0.001	✓(lin)	< 0.001	✓(k5)	< 0.001	✓(lin)	< 0.001	✓(lin)
No. positive lymph nodes	< 0.001	✓(k5)	< 0.001	✓(k5)	< 0.001	✓(k3)	< 0.001	✓(k5)	< 0.001	✓(k3)	< 0.001	✓(k3)
Hormonal therapy	0.867	✗	0.919	✗	0.504	✗	0.543	✗	0.576	✗	0.553	✗
Chemotherapy	< 0.001	✓(k5)	0.364	✓(k5)	< 0.001	✓(k5)	0.589	✓(k5)	< 0.001	✓(k5)	< 0.001	✓(k5)
Progesterone receptors	0.478	✗	0.600	✗	0.158	✗	0.117	✓(k5)	0.010	✗	0.119	✗

Abbreviations: MI, multiple imputation; SMC, substantive-model-compatible; TVE, time-varying effect.

Individuals were followed up from the time of breast cancer diagnosis to *disease recurrence* or *death due to breast cancer* (a composite event). Follow-up ranged from 1 to 231 months, and 1518 individuals (51%) had the outcome of interest. Following previous applications,^{12,42} we focus on the following eight covariates: age, tumour size 1 ($\leq 20\text{mm}$, $> 20\text{mm}$), tumour size 2 ($\leq 50\text{mm}$, $> 50\text{mm}$), tumour grade (grade 2 or 3 vs. grade 1), number of positive lymph nodes (transformed as $\exp(-2 \times 0.12 \times \text{no. positive lymph nodes})$), treatment with hormonal therapy (yes vs no), treatment with chemotherapy (yes vs. no), and progesterone receptors (pmol/l) (transformed as $\log(\text{progesterone receptors} + 1)$).

We generated missing data at random (MAR) in five covariates (tumour grade, number of positive lymph nodes, hormonal therapy, chemotherapy, and progesterone receptors) with the probability of missingness depending on age and tumour size (probability of missingness $e^{-9 + 0.1 \times \text{age} - \text{tumour size 2}} / (1 + e^{-9 + 0.1 \times \text{age} - \text{tumour size 2}})$). The missing data were generated conditionally independently (given age and tumour size 2) for each variable such that approximately 5% of individuals have missing data in any given variable. This resulted in 620 individuals (20% of the cohort) having missing data on at least one covariate.

A complete-data analysis was performed before missingness was introduced and a complete-case analysis was performed on the subset of 2362 individuals with no missing data. MI analyses were performed using the four methods considered in this paper: MI-Approx, MI-SMC, MI-TVE-Approx, and MI-TVE-SMC. All MI analyses used 20 imputations. All models considered included all eight covariates.

We first fitted a Cox regression model including TVEs for all variables, modelled using restricted cubic splines with 5 knots at the 5th, 25th, 50th, 75th, and 95th percentiles of the event times (0.51, 1.30, 2.54, 4.60, 9.12 years). Secondly, a test of the proportional hazards assumption was performed for each covariate using a joint Wald test (Section 4). Thirdly, we applied the MI-MTVE algorithm described in Section 4. The same procedure (steps 2-4 of the algorithm) was used in complete-data and complete-case analyses: in step 2, the complete-data or complete-case data set is used in place of the imputed data sets, and in step 3, standard joint Wald tests are used. A p-value cut-off of $\alpha = 0.01$ was used in the TVE model selection.

The results are shown in Figure 5, Supplementary Figure S9, and Table 2. On the basis of tests of the proportional hazard assumption, all analysis methods found evidence of TVEs for age, tumour size 1, tumour grade, number of positive lymph nodes, and chemotherapy (Table 2). There was little evidence of a TVE for tumour size 2, hormonal therapy, or progesterone receptors, except for progesterone receptors using MI-Approx. The form of TVEs was found to be broadly similar across all methods (Figure 5, Supplementary Figure S9). The complete-data and complete-case analyses gave very similar results. This is not unexpected here because the data were generated to be missing dependent only on covariates and not the outcome. Results from MI-Approx, MI-TVE-Approx, and MI-SMC-TVE were almost identical, whereas results from MI-SMC tended to differ from results of the other MI analyses. The confidence intervals obtained using our favoured imputation approach, MI-TVE-SMC (Figure 5), are not very noticeably narrower than from the complete-case analysis in this example, though in general, as found in the simulation study, the MI analysis will be more efficient. Older age was associated with an increasingly negative association with the hazard over time. Larger tumour size was associated with an increased hazard at the start of follow-up, which reduced over time. Those with tumour grade 2 or 3 had a much increased hazard relative to those with tumour grade 1 up to around 3 years of follow-up, after which a higher tumour grade was associated with a slightly reduced hazard. A higher number of positive lymph nodes was associated with higher hazard (noting that the variable is transformed in the analyses), and the association became stronger with increased follow-up. At the start of follow-up, chemotherapy was associated with an increased hazard but in later follow-up was associated with an increasingly reduced hazard.

The MI-MTVE algorithm selected TVEs for the same variables using all methods (Table 2), with the exception of tumour size 2, which was selected to have a TVE using MI-Approx, MI-TVE-Approx and MI-TVE-SMC but not the other methods and progesterone receptors, which was selected to have a TVE by MI-SMC. The functional forms chosen by the algorithm differed somewhat across methods.

In summary, similar conclusions would be reached from all analyses in this example. Further investigations looking at different missing data mechanisms and using different example data sets would be of interest. Sauerbrei et al^{12,43} applied the MFPT variable selection procedure on the basis of fractional polynomial models to this data set (using the complete data), identifying TVEs for tumour size 1 and progesterone receptors. Further work on how best to perform variable selection incorporating TVEs is needed even without the complication of missing data.

7 | DISCUSSION

In this article, we have introduced two multiple imputation methods allowing for TVEs to be included in Cox regression models. In the absence of TVEs, the methods of White and Royston⁵ (MI-Approx) and Bartlett et al⁶ (MI-SMC) can be used. MI-Approx is conceptually simpler, more convenient to code, and faster to run, whereas MI-SMC method has better

statistical properties. Our two proposals are extensions of these methods. The methods were described for a general functional form for TVEs. Researchers use different approaches to modelling TVEs. The correct functional form for a TVE is typically not known in advance, and thus, it is desirable to allow a flexible form. We therefore focused on a situation in which TVEs are modelled using restricted cubic splines. In some studies, it may be desirable for the TVE to be a simple step function, and we provided details on this in the Supplementary Materials (Section S3). In simulation studies, we used imputation models assuming a five-knot restricted cubic spline for the TVE. For binary covariates with missing data, both of our proposed methods performed well. However, the performance of the approximate method (MI-TVE-Approx) was slightly disappointing for continuous covariates when the effect size is large, though it still outperformed the complete-case analysis in all but scenario 2, and the observed bias was found to be smaller when the proportion with missing data is lower. The SMC method (MI-TVE-SMC) performed well across all scenarios, minimising bias, retaining approximately correct type I error rates in tests for nonproportional hazards when there is no TVE, and maximising power to detect TVEs when they exist compared with other MI methods. Our results showed that ignoring TVEs in the imputation model will result in incorrect type I errors (that is, not equal to the nominal level α) in the test for nonproportional hazards when the null hypothesis of proportional hazards is true and in a large loss of power to detect a TVE when one exists.

In practice, TVEs will often appear in the context of model building, including tests of the proportional hazards assumption. We therefore proposed the MI-MTVE model selection algorithm, an adaptation the MFPT algorithm of Sauerbrei et al,¹² for such settings. There remains work to be done to assess the performance of the MI-MTVE model selection algorithm and to compare it with possible alternatives. Our procedures for testing the proportional hazard assumption and the MI-MTVE algorithm draw on earlier work by Wood et al³³ on variable selection methods using multiply imputed data. There is a sizeable literature on model selection incorporating estimation of TVEs without including treatment of missing data. Berger et al⁴⁴ proposed the use of fractional polynomials⁴⁵ to select parsimonious forms for TVEs in Cox regression. Sauerbrei et al¹² proposed a model selection algorithm for use in Cox regression in which both the functional form for continuous covariates and the functional form for TVEs of covariates are modelled using fractional polynomials (the MFPT algorithm). See also chapter 11 in Royston and Sauerbrei.⁴² Abrahamowicz et al¹⁰ and Wynant and Abrahamowicz¹¹ also described methods for joint estimation of time-varying and nonlinear effects based on splines. Areas for further work include the extension of the methods proposed in this paper to a setting in which TVEs are modelled using fractional polynomials and to allow selection of functional forms for continuous variables and covariate interactions. This will build on the work Morris et al³⁴ on how to incorporate MI into a fractional polynomial model building procedure for explanatory variables. The MI-TVE-SMC approach is particularly suitable for extensions involving transformed covariates. Finally, further work is needed to investigate the validity of inferences following data-dependent model selection processes in the missing data context.

We have provided R code (<https://github.com/ruthkeogh/MI-TVE>), which can be used to implement the proposed imputation methods and perform analyses of the example data set used in Section 6. Moreover, MI-TVE-Approx is straightforward to apply in standard software, and although we provide example R code, this method can also be easily applied in Stata (`mi impute`) or SAS (`PROC MI`), for example. In addition, MI-SMC (not incorporating TVEs) can be applied using the `smcfcs` packages in R and Stata.⁴⁶ We have also provided an adaptation of this code for implementing MI-TVE-SMC in R. We applied our proposed methods to the analysis of the Rotterdam breast cancer study data, followed by the MI-MTVE model selection algorithm. The methods led to similar conclusions in this case. The data set is freely available, and we provide code at the above webpage, which enables the user to verify the analysis and change it for their own purposes.

There are of course limitations to this work. In particular, in some settings, complete-case analysis will be unbiased and MI biased. It follows that our methods are only applicable to settings in which MI is judged to be the best approach. When using our proposed methods, other forms of misspecification of the imputation model could result in bias, and this should be borne in mind, as in any MI analysis, especially for partially missing variables that are continuous, for which the normality assumption may not hold. Further work is also needed to investigate the performance of different approaches to model selection in this context. The general results given for MI-TVE-Approx and MI-TVE-SMC assumed that any censoring occurs independently of covariates with missing data. In MI-TVE-Approx, censoring depending on covariates with missing data can be accommodated by adding a further term, $\hat{H}_C(T)$, into the imputation model, which denotes the Nelson-Aalen estimate of the cumulative hazard for the censoring.¹⁷ In addition, MI-SMC has been extended to allow competing risks,⁴⁷ and this can be used to handle dependence of right-censoring on variables with missing data by modelling the censoring as a competing event. Moreover, MI-TVE-SMC can be extended in the same way. In both cases, it is assumed that the association between covariates and the hazard for censoring is not time-varying. Event times are also commonly subject to left truncation. Using MI-TVE-Approx, it can be shown that left truncation can be accommodated by

replacing $\hat{H}(T)$ by $\hat{H}(T) - \hat{H}(T_L)$ (and $\hat{H}^{(1)}(T)$ by $\hat{H}^{(1)}(T) - \hat{H}^{(1)}(T_L)$) in the imputation model and $\hat{H}_C(T)$ by $\hat{H}_C(T) - \hat{H}_C(T_L)$ if the censoring is suspected to depend on partially missing covariates. MI-TVE-SMC can also be extended to accommodate left truncation; however, further work is needed on this topic, including to implement the methods in the software.

We have focused on estimation of TVEs by modelling these in the Cox regression model. There exist other methods for estimating and testing for TVEs in Cox regression. Schoenfeld residual plots can be used to visually assess the proportional hazards assumption,⁴⁸ and smoothed residuals can be used to estimate TVEs.^{49,50} Scheike and Martinussen⁵¹ outlined tests for proportional hazards based on an iterative procedure to estimate cumulative regression coefficients, which does not require specification of the functional form for TVEs. Ng'andu⁵² summarized and compared several tests for the proportional hazards assumption. Other methods for estimating TVEs include those based on a kernel-weighted local partial likelihood⁵³ and penalized partial likelihoods.⁵⁴⁻⁵⁶ Gerds and Schumacher⁵⁷ proposed a measure for use in choosing between different models for the TVE to discover which is closest to the true shape.

In summary, for settings in which MI is judged to be appropriate and TVEs are a feature of the analysis, the approaches we have described should be used. Where computational time is not too large, the MI-TVE-SMC approach is recommended, though MI-TVE-Approx should also perform well if all covariates with missingness are binary or if effect sizes are small. Ignoring TVEs in the imputation may result in biased estimates and misleading conclusions.

ACKNOWLEDGEMENTS

The authors are grateful to Professor Ian White (MRC Clinical Trials Unit at UCL, UK), Professor Mike Kenward, and Dr Jonathan Bartlett (Department of Mathematical Sciences, University of Bath, UK) for comments on this work and to Professor Patrick Royston (MRC Clinical Trials Unit at UCL, UK) for advice on the example.

Ruth Keogh is funded by a Medical Research Council Methodology Fellowship (MR/M014827/1). Tim Morris is supported by the Medical Research Council (grant numbers MC_UU_12023/21 and MC_UU_12023/29).

ORCID

Ruth H. Keogh  <http://orcid.org/0000-0001-6504-3253>

Tim P. Morris  <http://orcid.org/0000-0001-5850-3610>

REFERENCES

1. Cox DR. Regression models and life tables. *J R Stat Soc Ser B*. 1972;34(2):187-202.
2. Cox DR. Partial likelihood. *Biometrika*. 1975;62(2):269-276.
3. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley & Sons; 1987.
4. Carpenter JR, Kenward MG. *Multiple Imputation and its Application*. Chichester, UK: John Wiley & Sons; 2013.
5. White IR, Royston P. Imputing missing covariate values for the Cox model. *Statist Med*. 2009;28(15):1982-1998.
6. Bartlett JW, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Stat Methods Med Res*. 2015;24(4):462-487.
7. Schemper M. Cox analysis of survival data with non-proportional hazard functions. *J R Stat Soc Ser D*. 1992;41(4):455-465.
8. Therneau TM, Grambsch PM. *Modelling Survival Data: Extending the Cox model*. New York, NY: Springer-Verlag; 2000.
9. Buchholz A, Sauerbrei W. Comparison of procedures to assess non-linear and time-varying effects in multivariable models for survival data. *Biom J*. 2011;53(2):308-331.
10. Abrahamowicz M, MacKenzie TA. Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Statist Med*. 2007;26(2):392-408.
11. Wynant W, Abrahamowicz M. Impact of the model-building strategy on inference about nonlinear and time-dependent covariate effects in survival analysis. *Statist Med*. 2014;33(19):3318-3337.
12. Sauerbrei W, Royston P, Look M. A new proposal for multivariable modelling of time-varying effects in survival data based on fractional-polynomial time-transformation. *Biom J*. 2007;49(3):453-473.
13. Binquet C, Abrahamowicz M, Mahboubi A, et al. Empirical study of the dependence of the results of multivariable flexible survival analyses on model selection strategy. *Statist Med*. 2008;27(30):6470-6488.
14. Seaman S, Galati J, Jackson D, Carlin J. What is meant by “missing at random”? *Stat Sci*. 2013;28(2):257-268.
15. Collett D. *Modelling Survival Data in Medical Research*. London, UK: Chapman and Hall; 1994.
16. Ford I, Norrie J. The role of covariates in estimating treatment effects and risk in long-term clinical trials. *Statist Med*. 2002;21(19):2899-2908. <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.1294>
17. Borgon O, Keogh R. Nested case-control: should one break the matching? *Lifetime Data Anal*. 2015;21(4):517-541.

18. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statist Med.* 2011;30(4):377-399.
19. Bartlett JW. *Correction for Classical Covariate Measurement Error and Extensions to Life-Course Studies* [PhD thesis]. London, UK: London School of Hygiene and Tropical Medicine; 2010.
20. Breslow NE. Discussion of the paper by D.R. Cox. *J R Stat Soc Ser B.* 1972;34:216-217.
21. Van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res.* 2007;16(3):219-242.
22. Quantin C, Abrahamowicz M, Moreau T, et al. Variation over time of the effects of prognostic factors in a population-based study of colon cancer: comparison of statistical methods. *Am J Epidemiol.* 1999;150(11):1188-1200.
23. Gore SM, Pocock SJ, Kerr GR. Regression models and non-proportional hazards in the analysis of breast cancer survival. *J R Stat Soc Ser C.* 1984;33(2):176-195.
24. Moreau T, O'Quigley J, Mesbah M. A global goodness-of-fit statistic for the proportional hazards model. *J R Stat Soc Ser C.* 1985;34(3):212-218.
25. Hess KR. Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statist Med.* 1994;13(10):1045-1062.
26. Heinzl H, Kaider A, Zlabinger G. Assessing interactions of binary time dependent covariates with time in Cox proportional hazards regression modelling using cubic spline functions. *Statist Med.* 1996;15(23):2589-2601.
27. Abrahamowicz M, MacKenzie T, Esdaile JM. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *J Am Stat Assoc.* 1996;91(436):1432-1439.
28. Hastie T, Tibshirani R. Varying-coefficient models. *J R Stat Soc Ser B.* 1993;55(4):757-796.
29. Gray RJ. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *J Am Stat Assoc.* 1992;87(420):942-951.
30. Kooperberg C, Stone CJ, Truong YK. Hazard regression. *J Am Stat Assoc.* 1995;90(429):78-94.
31. Stone C. Comment on Hastie and Tibshirani. *Stat Sci.* 1986;1:312-314.
32. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statist Med.* 1989;8(5):551-561.
33. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? *Statist Med.* 2008;27(17):3227-3246.
34. Morris T, White IR, Carpenter JR, Stanworth SJ, Royston P. Combining fractional polynomial model building with multiple imputation. *Statist Med.* 2015;34(25):32983317.
35. Li KH, Raghunathan TE, Rubin DB. Large-sample significance levels from multiply imputed data using moment-based statistics and an *F* reference distribution. *J Am Stat Assoc.* 1991;86(416):1065-1073.
36. Meng X-L, Rubin DB. Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika.* 1992;79(1):103-111.
37. Liu J, Gelman A, Hill J, Su Y-S, Kropko J. On the stationary distribution of iterative imputations. *Biometrika.* 2014;101(1):155-173.
38. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods.* 2001;6(4):330-351.
39. Buchholz A. *Assessment of Time-Varying Long-Term Effects of Therapies and Prognostic Factors* [PhD thesis]. Dortmund, Germany: Technical University of Dortmund; 2010.
40. Van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45(3):1-67.
41. Cummins DJ, Filloon TG, Nychka D. Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage. *J Am Stat Assoc.* 2001;96(453):233-246.
42. Royston P, Sauerbrei W. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Continuous Variables*. Chichester, UK: Wiley; 2008.
43. Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *J R Stat Soc Ser A.* 1999;162(1):71-94.
44. Berger U, Schäfer J, Ulm K. Dynamic Cox modelling based on fractional polynomials: time-variations in gastric cancer prognosis. *Statist Med.* 2003;22(7):1163-1180.
45. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat.* 1994;43(3):429-467.
46. Bartlett JW, Morris TP. Multiple imputation of covariates by substantive-model compatible fully conditional specification. *Stata J.* 2015;15:437-456.
47. Bartlett JW, Taylor JMG. Missing covariates in competing risks analysis. *Biostatistics.* 2016;17(4):751-763.
48. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika.* 1982;69(1):239-241.
49. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994;81(3):515-526.
50. Winnett A, Sasieni P. Iterated residuals and time-varying covariate effects in Cox regression. *J R Stat Soc Ser B.* 2003;65(2):473-488.
51. Scheike TH, Martinussen T. On estimation and tests of time-varying effects in the proportional hazards model. *Scand J Stat.* 2004;31(1):51-62.
52. Ng'andu NH. An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model. *Statist Med.* 1997;16(6):611-626.
53. Tian L, Zucker D, Wei LJ. On the Cox model with time-varying regression coefficients. *J Am Stat Assoc.* 2005;100(469):172-183.

54. Zucker DM, Karr AF. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann Stat*. 1990;18(1):329-353.
55. Verweij PJM, van Houwelingen HC. Time-dependent effects of fixed covariates in Cox regression. *Biometrics*. 1995;51(4):1550-1556.
56. Yan J, Huang J. Model selection for Cox models with time-varying coefficients. *Biometrics*. 2012;68(2):419-428.
57. Gerds TA, Schumacher M. Efron-type measures of prediction error for survival analysis. *Biometrics*. 2007;63(4):1283-1287. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-0420.2007.00832.x>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Keogh RH, Morris TP. Multiple imputation in Cox regression when there are time-varying effects of covariates. *Statistics in Medicine*. 2018;1–18. <https://doi.org/10.1002/sim.7842>