Survival analysis with time-dependent covariates subject to missing data or measurement error: Multiple Imputation for Joint Modeling (MIJM)

MARGARITA MORENO-BETANCUR*

Department of Epidemiology and Preventive Medicine, Monash University, 99 Commercial Rd, Melbourne, VIC 3004, Australia and Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute, 50 Flemington Rd, Parkville, VIC 3052, Australia

margarita.moreno@mcri.edu.au

JOHN B. CARLIN

Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute, Parkville, Australia and Melbourne School of Population and Global Health, University of Melbourne, Carlton, Australia

SAMUEL L. BRILLEMAN

Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia

STEPHANIE K. TANAMAS

Diabetes Epidemiology and Clinical Research Section, NIDDK, Phoenix, USA

ANNA PEETERS

School of Health and Social Development, Deakin University, Burwood, Australia

RORY WOLFE

Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia

SUMMARY

Modern epidemiological studies collect data on time-varying individual-specific characteristics, such as body mass index and blood pressure. Incorporation of such time-dependent covariates in time-to-event models is of great interest, but raises some challenges. Of specific concern are measurement error, and the non-synchronous updating of covariates across individuals, due for example to missing data. It is well known that in the presence of either of these issues the last observation carried forward (LOCF) approach traditionally used leads to bias. Joint models of longitudinal and time-to-event outcomes, developed

*To whom correspondence should be addressed.

recently, address these complexities by specifying a model for the joint distribution of all processes and are commonly fitted by maximum likelihood or Bayesian approaches. However, the adequate specification of the full joint distribution can be a challenging modeling task, especially with multiple longitudinal markers. In fact, most available software packages are unable to handle more than one marker and offer a restricted choice of survival models. We propose a two-stage approach, Multiple Imputation for Joint Modeling (MIJM), to incorporate multiple time-dependent continuous covariates in the semi-parametric Cox and additive hazard models. Assuming a primary focus on the time-to-event model, the MIJM approach handles the joint distribution of the markers using multiple imputation by chained equations, a computationally convenient procedure that is widely available in mainstream statistical software. We developed an R package "survtd" that allows MIJM and other approaches in this manuscript to be applied easily, with just one call to its main function. A simulation study showed that MIJM performs well across a wide range of scenarios in terms of bias and coverage probability, particularly compared with LOCF, simpler two-stage approaches, and a Bayesian joint model. The Framingham Heart Study is used to illustrate the approach.

1. Introduction

Large-scale studies in which data are collected from individuals over time are central to modern health research. Our motivating example and a pioneer of such studies in epidemiology is the Framingham Heart Study (FHS), in which a wide range of time-varying exposures have been collected regularly over a period of more than 55 years. These variables include both physical/biological characteristics [body mass index (BMI), blood pressure, etc.] and behavioral/socio-demographic characteristics (smoking, marital status, etc.). To understand the pathways by which such characteristics may affect the occurrence of health events (death, diabetes onset, etc.), it is common to incorporate them as time-dependent covariates in models for a time-to-event outcome.

Hazard models are often adopted, assuming that the hazard rate at a given time depends on the current "true" (error-free) values of the covariates. In principle, unbiased estimation of these models requires that we know, at each observed event time, the current true covariate values for the set of individuals still at risk just before that point. This requirement is rarely, if ever, fulfilled for time-dependent continuous covariates, or markers, for two main reasons. First, many markers, particularly biological markers, are measured with error due to flaws in technical procedures and biological, non-prognostic, and short-term variations. Second, these covariates are usually measured in discrete time, but changes in their values and failures both occur in continuous time. Thus, the true values of the markers at the event times can be viewed as missing data.

This missing data problem is commonly dealt with using the last observation carried forward (LOCF) approach, i.e. singly imputing a missing value at an event time with the last observation available. Andersen and Liestøl (2003) found that the bias imparted by this approach increases with the degree of measurement error and with non-synchronous updating of the covariates across individuals, for instance due to skipped follow-up visits. While missing visit-specific measurements are not those necessary to fit the model (i.e. those at the event times), this missingness exacerbates the biases arising from the discrete-time observation of continuously changing characteristics.

The caveats of LOCF have led to the development of joint modeling approaches, whereby the joint distribution of a number of time-dependent markers (usually just one) and the time-to-event outcome is modeled. Various maximum likelihood methods (e.g. De Gruttola and Tu, 1994; Wulfsohn and Tsiatis, 1997) and Bayesian approaches (e.g. Faucett and Thomas, 1996; Daniels and Hogan, 2008; Hanson *and others*, 2011; Rizopoulos, 2016b) have been developed to fit these models, and comprehensive reviews are

available (e.g. Wu *and others*, 2012). Drawbacks of these methods are the reliance of their performance on the correct specification of the full joint distribution, which can be a difficult task, and their computational complexity, which limits the flexibility of joint modeling packages available in mainstream statistical software. Indeed, most packages cannot handle more than one marker and are constrained to parametric hazard functions or proportional hazards models (Gould *and others*, 2015; Hickey *and others*, 2016). Restriction to such survival models is also evident in specific models proposed for the setting with multiple time-dependent markers (e.g. Xu and Zeger, 2001; Chi and Ibrahim, 2006; Rizopoulos and Ghosh, 2011).

Two-stage joint modeling approaches are an appealing alternative to full joint modeling to tackle multiple markers and other survival models, such as the additive hazard model. Various two-stage methods have been proposed for one marker (Wu *and others* 2012), but the general principle of the simpler versions is as follows. In the first stage, a linear mixed model is fitted to estimate the subject-specific true marker trajectories, and the missing values at the event times are singly imputed from these. In the second stage, the hazard model is fitted using the imputed data. These simple approaches rely on approximations to the marker distribution and can be less efficient than joint-estimation methods. Notably, issues could arise: (i) when the uncertainty in the single imputations is not accounted for in the second stage; (ii) when, with multiple markers, the interrelations among these are not accounted for in the first stage; and (iii) when the informative drop-out induced by the events is ignored in the first stage, leading to an inaccurate approximation of the marker distribution (Ye *and others*, 2008; Albert and Shih, 2010).

We propose a new two-stage approach, Multiple Imputation for Joint Modeling (MIJM), to incorporate multiple markers in the semi-parametric Cox and additive hazard models. We draw from the literature on multiple imputation (MI) (Rubin, 1987), a missing data approach, to improve upon simple two-stage approaches. The joint distribution of the markers is treated as a nuisance, which is computationally convenient, but also reasonable when interest lies in the time-to-event model and considering that adequately specifying that multivariate distribution may be challenging. The approach is easy to implement in mainstream software. The R package `survtd` (Moreno-Betancur, 2017) facilitates its application (see Section 7 of supplementary material available at *Biostatistics* online).

2. Notation and assumptions

Suppose that n individuals in a study are followed until an event of interest occurs (e.g. death) or a censoring time (e.g. due to end of study). We assume independence across individuals and independent censoring given covariates. For $i = 1, \ldots, n$, let T_i and C_i denote, respectively, the time to event and the right-censoring time, so the observed time is $\tilde{T}_i = \min\{T_i, C_i\}$ and the event indicator is $D_i = I(T_i < C_i)$, where I is the indicator function. Let $t_1, \ldots, t_S$ denote the set of times, indexed in increasing order, at which an event was observed (i.e. all $\tilde{T}_i$ such that $D_i = 1$). The aim is to model the hazard rate of the event, denoted by $\lambda(t)$ at time $t \geq 0$, as a function of a vector $\boldsymbol{Z}_i$ of time-fixed covariates measured at baseline, assumed to be fully-observed without error, and several, say K, time-varying continuous markers. Let $Y_i^k(t)$ denote the "true" (error-free) value of the kth marker for individual i at time t ($k = 1, \ldots, K$; $i = 1, \ldots, n$; $t \geq 0$).

2.1. *Target hazard rate models*

We consider two types of model, both assuming that the hazard rate depends on the current value of all markers and the time-fixed covariates, and leaving the baseline hazard, denoted by $\lambda_0(t)$, unspecified. The two models differ in the interpretation of their respective regression coefficients, providing complementary information of interest. The Cox proportional hazards model assumes that covariates have multiplicative effects on the hazard, in the following standard form:

$$\lambda\{t|\bar{\boldsymbol{Y}}_i(t), \boldsymbol{Z}_i\} = \lambda_0(t) \exp\{\boldsymbol{\gamma}'_Y \boldsymbol{Y}_i(t) + \boldsymbol{\gamma}'_Z \boldsymbol{Z}_i\}, \tag{2.1}$$

where $\boldsymbol{Y}_i(t) = \{Y_i^1(t), \dots, Y_i^K(t)\}'$ and $\bar{\boldsymbol{Y}}_i(t)$ denotes the history of the markers until time t. Regression coefficients in this model provide relative measures of association (hazard ratios). The semi-parametric additive hazard model (Lin and Ying, 1994) is the additive counterpart of model (2.1), and its coefficients provide absolute measures of association (hazard differences):

$$\lambda\{t|\bar{\boldsymbol{Y}}_i(t), \boldsymbol{Z}_i\} = \lambda_0(t) + \boldsymbol{\gamma}'_Y \boldsymbol{Y}_i(t) + \boldsymbol{\gamma}'_Z \boldsymbol{Z}_i. \tag{2.2}$$

2.2. *Assumptions about the marker measurement process*

In principle, unbiased estimation of models (2.1) and (2.2) requires data on the true marker values at the observed event times, but these are virtually never available. The trajectories $\{Y_i^k(t) : t \geq 0\}$ are not observed in continuous time, but are often to be measured at a set of pre-fixed visit times, $\tau_1 := 0, \tau_2, \tau_3, \dots$, common to all individuals, until the event or censoring occurs. The visit times will not generally coincide with the event times $t_1, \dots, t_S$. Let J_i denote the number of planned visit times occurring before $\tilde{T}_i$. Further, for each marker some of the J_i measurements can be missing (e.g. due to skipped visits). We refer to these as missing "at-risk-visit" measurements, to distinguish them both from the missing values at the event times needed to fit the hazard models and from potentially missing measurements after the event or censoring occurs. We assume that the probability of a missing at-risk-visit value does not depend on the value itself given values of that marker at other visits, values of the other $K - 1$ markers at that visit, the time-fixed covariates, the time-to-event outcome and visit timing. This is a more relaxed condition than the "everywhere missing at random" assumption (Seaman *and others*, 2013).

Finally, when a measurement does take place, the value is measured with random error, represented as a normally distributed deviation from the true trajectory. Hence, for k and i fixed, the available measurements are $\tilde{Y}_{ij}^k$, for j in a subset of $\{1, \dots J_i\}$, where $\tilde{Y}_{ij}^k = Y_i^k(\tau_j) + \epsilon_{ij}^k$ and $\epsilon_{ij}^k \sim N(0, \sigma^{2,k})$ is the measurement error. The deviations ϵ_{ij}^k are assumed to be independent of the covariates, the markers and of each other across time-points (j) and markers (k).

3. Multiple Imputation for Joint Modeling (MIJM)

3.1. *Overview of two-stage joint modeling approaches*

The data missing to fit models (2.1) and (2.2) are the values $Y_i^k(t_s)$ for $k = 1, \dots, K$, $i = 1, \dots, n$ and $s \in \{1, \dots, S\}$ such that $t_s \leq \tilde{T}_i$. Two-stage approaches generally seek to impute these with probable values based on fitted models (first stage), and then use these imputations to fit the hazard model (second stage). Note that there is no need to impute the missing at-risk-visit values.

A simple version is to fit separately, in the first stage, K linear mixed models, one per marker, including as predictors the covariates $\boldsymbol{Z}_i$ and spline or polynomial terms for functions of time (and possibly interaction terms). This yields smoothed estimates of the subject-specific true marker trajectories from which the missing values are singly imputed. The performance of this approach (henceforth "simple2S") can be suboptimal due to issues (i)–(iii) mentioned in the introduction.

The proposed MIJM approach, described in the following sections, is a refined version using MI, under which multiple, say M, imputations of the missing values are drawn from interdependent imputation models (in this case one model for each marker) to produce M completed data sets. The analysis model (in this case the hazard model) is fitted to each completed data set and the M estimates obtained from these are then combined using the formulas of Rubin (1987). The benefit of multiple over single imputation is that MI standard errors appropriately reflect the uncertainty regarding the imputed values, addressing the aforementioned issue (i) of simple two-stage approaches

(Bycott and Taylor, 1998; Ye *and others*, 2008; Albert and Shih, 2010). Further, by carefully building the K interdependent imputation models, and drawing values from these using an iterative algorithm called multiple imputation by chained equations (MICE) (van Buuren and Groothuis-Oudshoorn, 2011), our MI approach also addresses issues (ii) and (iii).

3.2. *Stage 1: multiple imputation of time-dependent covariates*

3.2.1. *Imputation models for the longitudinal markers.* For each marker $k \in \{1, \ldots, K\}$, an imputation model is built by fitting a linear mixed model to the observed values of the marker across visits and individuals ($\tilde{Y}_{ij}^k$, $i = 1, \ldots, n$ and j in a subset of $\{1, \ldots J_i\}$). Specifically, expected subject-specific trajectories are modeled as:

$$(\boldsymbol{\beta}_Z^k)' \boldsymbol{Z}_i + (\boldsymbol{\beta}_Y^k)' \boldsymbol{Y}_i^{-k}(\tau_j) + (\boldsymbol{\beta}_\tau^k)' \boldsymbol{\eta}^k(\tau_j) + (\boldsymbol{\beta}_O^k)' \boldsymbol{O}_i(\tau_j) + (\boldsymbol{b}_i^k)' \boldsymbol{U}_i^k(\tau_j). \tag{3.1}$$

Components of this mixed model are now described in detail. The fixed-effects vector is $\boldsymbol{\beta}^k = \{(\boldsymbol{\beta}_Z^k)', (\boldsymbol{\beta}_Y^k)', (\boldsymbol{\beta}_\tau^k)', (\boldsymbol{\beta}_O^k)'\}'$ and the random-effects vectors are $\boldsymbol{b}_i^k \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^k)$, $i = 1, \ldots, n$. The vector $\boldsymbol{\eta}^k(\tau_j)$ may contain spline or polynomial terms for τ_j to model time-trends in the fixed-effects part of the model, which also includes the time-fixed covariates $\boldsymbol{Z}_i$. A subset of the variables in $\boldsymbol{Z}_i$ and $\boldsymbol{\eta}^k(\tau_j)$, collected in the vector $\boldsymbol{U}_i^k(\tau_j)$, are allowed to have subject-specific random effects. Such flexibility may be needed for markers with complex-shaped trajectories. In our simulation study and example, we considered random intercepts and slopes. Another two key sets of variables are included as predictors with fixed effects in (3.1), as described next.

First, the full vector of current values of the other $K - 1$ markers, $\boldsymbol{Y}_i^{-k}(\tau_j) = \{Y_i^1(\tau_j), \ldots, Y_i^{k-1}(\tau_j), Y_i^{k+1}(\tau_j), \ldots, Y_i^K(\tau_j)\}'$, is included. All these variables are in the target hazard model and could also be predictive of missingness, and recommendations in the MI literature suggest including all variables in any of these two categories in the imputation model (van Buuren and Groothuis-Oudshoorn, 2011). This can also be seen as an approximate way to deal with issue (ii) of simple two-stage approaches, i.e. the interrelations among the markers are accounted for.

Second, information on the time-to-event outcome is included following recommendations for MI in the context of incomplete covariates (van Buuren *and others*, 1999). This addresses issue (iii) in that it provides an approximation of the marker distribution that can be sufficiently accurate for the purpose of imputation and subsequent estimation of the target time-to-event model. Specifically, we include: the Nelson–Aalen estimator of cumulative hazard at time $\tilde{T}_i$, $\widehat{\Lambda}(\tilde{T}_i)$; interactions of $\widehat{\Lambda}(\tilde{T}_i)$ with the time-fixed covariates (succinctly, $\widehat{\Lambda}(\tilde{T}_i) \times \boldsymbol{Z}_i$); and a modified event indicator, defined by $D_{ij} = 1$ if $D_i = 1$ and $j = J_i$, and $D_{ij} = 0$ otherwise, i.e. $D_{ij} = I(D_i = 1 \wedge J_i = j)$. Thus, in (3.1), $\boldsymbol{O}_i(\tau_j)$ is the vector containing $\widehat{\Lambda}(\tilde{T}_i)$, $\widehat{\Lambda}(\tilde{T}_i) \times \boldsymbol{Z}_i$ and D_{ij}. The justification for this strategy, used by the MIJM approach, is provided in the Supplementary material available at *Biostatistics* online (Section 1) and extends results of White and Royston (2009). Briefly, it is based on an approximation to the conditional distribution of each marker given the time-to-event outcome.

Throughout the manuscript, we also consider an alternative approach, identical to MIJM in all aspects except that it includes the actual event indicator D_i as predictor in the imputation model instead of the modified indicator, i.e. $\boldsymbol{O}_i(\tau_j)$ contains $\widehat{\Lambda}(\tilde{T}_i)$, $\widehat{\Lambda}(\tilde{T}_i) \times \boldsymbol{Z}_i$, and D_i. This is the approach recommended to impute incomplete time-fixed covariates for the Cox model (White and Royston, 2009), and we assessed its comparative performance here due to its simplicity. Nonetheless, we expect that approximation to be less accurate as it is unadapted to the time-dependent covariate setting, and we thus refer to it as "unMIJM". For unMIJM, $\boldsymbol{O}_i(\tau_j) = \boldsymbol{O}_i(\tau_{j'})$ for all j, j', i.e. the vector is time-fixed, but we keep the same notation throughout for simplicity.

As usual with MI, the resulting K imputation models are not assumed necessarily to represent the true data generating mechanism (because of the difficulty of specifying a joint distribution that is compatible with these conditional models) but are working models for imputation purposes.

3.2.2. *Multiple imputation by chained equations (MICE).* The drawing of imputations for the missing marker values at the event times (i.e. $Y_i^k(t_s)$ for s such that $t_s \leq \tilde{T}_i$) from the trajectories determined by (3.1) requires an iterative procedure due to the dependence of each marker model on the true values of other markers, which need to be imputed as well. We use MICE, a procedure that draws iteratively from these K models until convergence to yield one set of imputations.

For $k = 1, \ldots, K$, let $\boldsymbol{\Theta}^k = \{(\boldsymbol{\beta}^k)', (\boldsymbol{b}_1^k)', \ldots, (\boldsymbol{b}_n^k)'\}'$, and let $\tilde{\boldsymbol{Y}}_{obs}^k$ denote the vector of all the available measurements of marker k across individuals. Let $\boldsymbol{Y}^k$ collect the true values of this marker across all individuals at their respective relevant planned visit and event times (for individual i, the values for each $t \in \{\tau_1, \ldots, \tau_{J_i}, t_1, \ldots, t_S\}$ such that $t \leq \tilde{T}_i$). Let $\boldsymbol{\eta}_i^k = \{\boldsymbol{\eta}^k(t)' : t = \tau_1, \ldots, \tau_{J_i}, t_1, \ldots, t_S\}'$ and $\boldsymbol{U}_i^k = \{\boldsymbol{U}_i^k(t)' : t = \tau_1, \ldots, \tau_{J_i}, t_1, \ldots, t_S\}'$. For the MIJM method, we extend the definition of $\boldsymbol{O}_i(t)$ to any time-point $t \notin \tau_1, \tau_2, \ldots$ by considering the natural definition of the modified indicator at time t: $D_i(t) = I(D_i = 1 \wedge \tilde{T}_i = t)$ (see Section 1 of supplementary material available at *Biostatistics* online) and define $\boldsymbol{O}_i = \{\boldsymbol{O}_i(t)' : t = \tau_1, \ldots, \tau_{J_i}, t_1, \ldots, t_S\}'$. For unMIJM, we let $\boldsymbol{O}_i(t) = \boldsymbol{O}_i$ denote the corresponding time-constant vector. The vectors $\boldsymbol{X}_i^k = \{\boldsymbol{Z}_i', (\boldsymbol{\eta}_i^k)', \boldsymbol{O}_i', (\boldsymbol{U}_i^k)'\}'$, $i = 1, \ldots, n$, are collected in $\boldsymbol{X}^k$.

For imputation $m \in \{1, \ldots, M\}$, the iterative procedure is initialized by imputations based on simple random draws $\boldsymbol{Y}_{(0)}^k$, $k = 1, \ldots, K$, from the data. If G iterations are to be carried out, then at iteration $g \in \{1, \ldots, G\}$, draws $\boldsymbol{\Theta}_{(g)}^k$ and $\boldsymbol{Y}_{(g)}^k$ of $\boldsymbol{\Theta}^k$ and $\boldsymbol{Y}^k$ are performed sequentially for $k = 1, \ldots, K$ as follows, based on the current values of the remaining markers:

$$\boldsymbol{\Theta}_{(g)}^k \sim P(\boldsymbol{\Theta}^k | \tilde{\boldsymbol{Y}}_{obs}^k, \boldsymbol{Y}_{(g)}^1, \ldots, \boldsymbol{Y}_{(g)}^{k-1}, \boldsymbol{Y}_{(g-1)}^{k+1}, \ldots, \boldsymbol{Y}_{(g-1)}^K, \boldsymbol{X}^k) \text{ and}$$

$$\boldsymbol{Y}_{(g)}^k \sim P(\boldsymbol{Y}^k | \tilde{\boldsymbol{Y}}_{obs}^k, \boldsymbol{Y}_{(g)}^1, \ldots, \boldsymbol{Y}_{(g)}^{k-1}, \boldsymbol{Y}_{(g-1)}^{k+1}, \ldots, \boldsymbol{Y}_{(g-1)}^K, \boldsymbol{X}^k, \boldsymbol{\Theta}_{(g)}^k),$$

where the formulations for $k = 1$ and $k = K$ make the obvious modification. These draws are performed following *(I)* and *(II)* below, which are based on the procedure of Moreno-Betancur and Chavance (2016) but incorporating measurement error correction.

(I) Performing draws of the parameters $\boldsymbol{\Theta}_{(g)}^k$.

Using standard asymptotic results, the conditional (posterior) distribution of $\boldsymbol{\Theta}_{(g)}^k$ above is approximated by a normal distribution with mean and variance equal to the estimates of fixed and random effects and of their variance–covariance, respectively, given restricted maximum likelihood (REML) estimates of the variance components. Specifically, the imputation model for marker k is fitted based on the current values of the other markers using the profiled REML criterion of Bates *and others* (2015). This yields maximum likelihood estimates and best linear unbiased predictors of the fixed and random effects vectors, respectively, denoted by $\hat{\boldsymbol{\beta}}_{(g)}^k$ and $\hat{\boldsymbol{b}}_{i,(g)}^k$, $i = 1, \ldots, n$, and estimates of their variance–covariance matrices, denoted by $\hat{\boldsymbol{V}}_{\boldsymbol{\beta},(g)}^k$ and $\hat{\boldsymbol{V}}_{\boldsymbol{b}_i,(g)}^k$ $i = 1, \ldots, n$. We thus draw $\boldsymbol{\beta}_{(g)}^k \sim N\{\hat{\boldsymbol{\beta}}_{(g)}^k, \hat{\boldsymbol{V}}_{\boldsymbol{\beta},(g)}^k\}$ and $\boldsymbol{b}_{i,(g)}^k \sim N\{\hat{\boldsymbol{b}}_{i,(g)}^k, \hat{\boldsymbol{V}}_{\boldsymbol{b}_i,(g)}^k\}$ for $i = 1, \ldots, n$, and set $\boldsymbol{\Theta}_{(g)}^k = \{(\boldsymbol{\beta}_{(g)}^k)', (\boldsymbol{b}_{1,(g)}^k)', \ldots, (\boldsymbol{b}_{n,(g)}^k)'\}'$.

(II) Performing draws of $\boldsymbol{Y}_{(g)}^k$ from the posited trajectory.

Draws of marker k for individual i are required for all times $t \in \{\tau_1, \ldots, \tau_{J_i}, t_1, \ldots, t_S\}$ such that $t \leq \tilde{T}_i$. Given $\boldsymbol{\Theta}_{(g)}^k$, a draw at time t is obtained using the linear predictor (3.1), setting $Y_{i,(g)}^k(t) = (\boldsymbol{\beta}_{Z,(g)}^k)'\boldsymbol{Z}_i + (\boldsymbol{\beta}_{Y,(g)}^k)'\boldsymbol{Y}_{i,(g)}^{-k}(t) + (\boldsymbol{\beta}_{\tau,(g)}^k)'\boldsymbol{\eta}^k(t) + (\boldsymbol{\beta}_{O,(g)}^k)'\boldsymbol{O}_i(t) + (\boldsymbol{b}_{i,(g)}^k)'\boldsymbol{U}_i^k(t)$, where $\boldsymbol{Y}_{i,(g)}^{-k}(t) = \{Y_{i,(g)}^1(t),$

$\dots, Y_{i,(g)}^{k-1}(t), Y_{i,(g-1)}^{k+1}(t), \dots, Y_{i,(g-1)}^{K}(t)\}'$ is the vector of current values of the other markers at time t for individual i.

At convergence, often achieved after a small number of iterations, the algorithm yields the mth set of imputations. In practice, the number of iterations should be based on graphical assessments of convergence (van Buuren and Groothuis-Oudshoorn, 2011), while the number of imputations may be based on relative efficiency calculations (Rubin, 1987). Note that, in addition to imputing the missing values at the observed event times $t_1, \dots, t_S$ at which the individual is at risk, imputations are also performed at each iteration for all planned visit times $\tau_1, \tau_2, \dots$ even though these values are not needed to fit the hazard model. The reason is that the imputation model (3.1) is estimated at each iteration from the values occurring at planned visit times. In practical terms, this requires stacking two distinct data sets for the imputation procedure.

3.3. *Stage 2: fitting the hazard model*

The procedure above yields M imputations of each missing value. To fit model (2.1) or (2.2) based on the mth set of imputations ($m = 1, \dots, M$), we require a data set with multiple rows per individual, one for each event-time t_s such that $t_s \leq \tilde{T}_i$. The data set has a column for each of the K markers, including its mth imputed value for the corresponding individual and event time. Additional columns include the time-fixed covariates $\boldsymbol{Z}_i$, which are constant across rows corresponding to the same individual. Columns defining the entry and exit times, as in the usual data set-up when dealing with time-dependent covariates, are also required (Therneau and Grambsch, 2000). Thus, for the row corresponding to individual i and event time t_s, the "entry" time is the previous event-time t_{s-1} (where we define $t_0 = 0$) and the "exit" time is t_s. The data set prepared in this way can then be used in a function that fits the hazard model using a start/stop or entry/exit formulation. The M results obtained are then combined using Rubin's formulas. The main function of the R `survtd` package (Moreno-Betancur, 2017) performs all the steps: data manipulation pre- and post-MI, MI procedure, model fitting, and pooling of results, so that one call to this function suffices to apply our approach (see Section 7 of supplementary material available at *Biostatistics* online).

4. SIMULATION STUDY

We aimed to compare the performance across various scenarios of the following approaches: LOCF, simple2S, unMIJM, and MIJM. We also considered a Bayesian full joint modeling method ("JMbayes"), available in a development version of the R package `JMbayes` (Rizopoulos, 2016a), but only for the Cox model in a subset of scenarios. This method assumes a multivariate mixed model for the markers and a hazard model similar to (2.1) but the baseline hazard is modeled using B-splines (Rizopoulos, 2016b).

4.1. *Data generation*

The simulation models closely followed aspects of the FHS, but we considered a reduced sample size ($n = 200$, except when assessing impact of sample size) and a less intensive and shorter schedule of planned visit times relative to the median survival time as not many studies have the luxury of such a long and intensive follow-up. For each individual $i = 1, \dots, n$, we generated two time-fixed covariates: $Z_{1i} \sim \mathcal{B}ern(0.45)$ and $Z_{2i} \sim N(44, 8.5^2)$, following the distribution of gender and baseline age, respectively, in the FHS. Parameters of the models used to generate the time-dependent markers, missing data, and the time-to-event outcome, described next, were determined from the FHS by fitting analogous models to, respectively, systolic blood pressure (SBP) measurements, the probability of missing at-risk-visit SBP measurements, and mortality. Some parameters were fixed while others were varied across simulation scenarios (see next section). All parameter values and related details are provided in the Supplementary material available at *Biostatistics* online (Section 2).

Repeated measures of three markers ($K = 3$) were generated for the n subjects at common times $\tau_j \in \{0, 1, \ldots, 9\}$ such that $\tilde{Y}_{ij}^k = Y_i^k(\tau_j) + \epsilon_{ij}^k$, where $\epsilon_{ij}^k \sim N(0, \sigma^{2,k})$ were independent across j and k. We considered two types of model, detailed next, for the joint distribution of the multiple markers to determine the true underlying trajectories. Fixed intercept parameters were chosen in both cases so that the marginal mean of the error-polluted measurements was around 140 in all scenarios, as is the case for SBP (in mmHg) in the FHS. The marginal standard deviation varied between 30 and 50 according to the measurement error scenario (see below).

The first was a correlated random effects model, which is a multivariate mixed model, parametrizing the marginal mean of the marker vector (Fieuws and Verbeke, 2004). We drew from trajectories given by $Y_i^k(\tau_j) = \alpha_0^k + \alpha_{Z_1}^k Z_{1i} + \alpha_{Z_2}^k Z_{2i} + \alpha_\tau^k \tau_j + a_{i0}^k + a_{i\tau}^k \tau_j$, for $k = 1, 2, 3$, where $\boldsymbol{a}_{i0} = (a_{i0}^1, a_{i0}^2, a_{i0}^3)$ and $\boldsymbol{a}_{i\tau} = (a_{i\tau}^1, a_{i\tau}^2, a_{i\tau}^3)$ were both zero-mean multivariate normal vectors, for $i = 1, \ldots, n$, such that $\text{corr}(a_{i0}^k, a_{i0}^l) = \text{corr}(a_{i\tau}^k, a_{i\tau}^l) = \rho \geq 0$, for all $k, l \in \{1, 2, 3\}$. Hence, the random intercepts across markers could be correlated, and so could the random slopes, and this correlation ρ modulated the correlation between markers. The random intercepts and slopes were independent of each other, i.e. $\text{corr}(a_{i0}^k, a_{i\tau}^l) = 0$, for all $k, l \in \{1, 2, 3\}$. Hence $\boldsymbol{a}_{i0}^k = (a_{i0}^k, a_{i\tau}^k) \sim N(\boldsymbol{0}, \boldsymbol{\Omega}^k)$ where $\boldsymbol{\Omega}^k$ was a diagonal matrix with diagonal $(\omega_0^{2,k}, \omega_\tau^{2,k})$.

The second was the so-called "product-normal" model, which is based on conditional models (Spiegelhalter, 1998; Cooper *and others*, 2007). Here, the trajectory of marker 1 did not depend on any of the other markers and the trajectory of marker $k \geq 2$ depended on the current values of markers 1 through $k - 1$. Thus the markers were generated sequentially from a chain of conditional models, assuming that the trajectory of marker $k \in \{1, 2, 3\}$ followed $Y_i^k(\tau_j) = \alpha_0^k + \alpha_{Z_1}^k Z_{1i} + \alpha_{Z_2}^k Z_{2i} + \sum_{l=1}^{k-1} \alpha_{Y_l}^k Y_i^l(\tau_j) + \alpha_\tau^k \tau_j + a_{i0}^k + a_{i\tau}^k \tau_j$, where the fourth term was null for $k = 1$ and $\boldsymbol{a}_i^k = (a_{i0}^k, a_{i\tau}^k)' \sim N(\boldsymbol{0}, \boldsymbol{\Omega}^k)$, $i = 1, \ldots, n$, were vectors of subject-specific random intercepts and slopes independent of each other (i.e. with $\boldsymbol{\Omega}^k$ diagonal as for the correlated random effects model) and independent of the random effects for other markers. Here the interdependence among markers was controlled through the regression coefficients $\alpha_{Y_l}^k$.

For both the Cox and additive hazard models, a time to event T_i was generated for each individual using the cumulative hazard inversion method, assuming a Weibull baseline hazard with shape parameter $\upsilon = 2$ and scale parameter λ. For the Cox model, the hazard function was given by: $\lambda\{t | Z_{1i}, Z_{2i}, \bar{Y}_i^1(t), \bar{Y}_i^2(t), \bar{Y}_i^3(t)\} = 2\lambda t \exp\left\{\gamma_{Z_1} Z_{1i} + \gamma_{Z_2} Z_{2i} + \sum_{k=1}^3 \gamma_{Y_k} Y_i^k(t)\right\}$, and for the additive hazard model by $\lambda\{t | Z_{1i}, Z_{2i}, \bar{Y}_i^1(t), \bar{Y}_i^2(t), \bar{Y}_i^3(t)\} = 2\lambda t + \gamma_{Z_1} Z_{1i} + \gamma_{Z_2} Z_{2i} + \sum_{k=1}^3 \gamma_{Y_k} Y_i^k(t)$. In both cases around 20% independent censoring was superimposed by drawing censoring times C_i from a uniform distribution. The repeated measures of the K markers were then truncated at the follow-up time $\tilde{T}_i = \min\{T_i, C_i\}$, so only $J_i = \max\{j : \tau_j \leq \tilde{T}_i\}$ measurements were available for individual i at this stage. Then measurement $\tilde{Y}_{ij}^k, 0 < j \leq J_i$, was set to missing if $R_{ij}^k = 1$ where R_{ij}^k was drawn from a Bernoulli distribution with success probability following $\text{logit}\{P(R_{ij}^k = 1)\} = \delta_0^k + \delta_\tau^k \tau_j + \delta_T^k \tilde{T}_i + \delta_D^k D_i + \delta_{Z_1}^k Z_{1i} + \delta_{Z_2}^k Z_{2i} + \sum_{l \neq k} \delta_l^k Y_i^l(\tau_j)$.

4.2. *Simulation plan*

In preliminary simulations, we assessed the procedure in the case of one marker ($K = 1$). With $K = 3$ markers, we performed a main set of simulations for each combination of multiple marker model and hazard rate model, altering the following aspects for each marker $k = 1, 2, 3$ to ranges observed in the FHS example. First, we varied $\sigma^{2,k}$ to alter the degree of measurement error, using the following intra-cluster correlation (ICC)-type measure as an indicator of the scale of error relative to total variation: $\text{ICC}^k = (\omega_0^{2,k} + \omega_\tau^{2,k})/(\omega_0^{2,k} + \omega_\tau^{2,k} + \sigma^{2,k})$. The lower the measurement error, the higher ICC^k. We considered scenarios with low ($\text{ICC}^k = 0.8$), moderate ($\text{ICC}^k = 0.5$) and high ($\text{ICC}^k = 0.2$) error. Second, we varied δ_0^k to alter the proportion p_{Miss}^k of missing at-risk-visit values to no ($p_{\text{Miss}}^k = 0\%$), low ($p_{\text{Miss}}^k \approx 20\%$), and high ($p_{\text{Miss}}^k \approx 40\%$) missingness.

Third, we considered null, weak and strong effects of the markers on the hazard by taking $\gamma_{Y_k} \in \{0, 0.005, 0.01\}$ in the Cox model and $\gamma_{Y_k} \in \{0, 0.0005, 0.001\}$ in the additive model, respectively. Estimates of these parameters were expressed as effects per 100 units for the Cox model and per 1000 units for the additive model. Fourth, we varied the strength of the pairwise marginal correlations between the three markers to low (around 0.05), moderate (around 0.15), or high (around 0.3), by setting, respectively, $\rho = 0, 0.5, 0.75$ in the correlated random effects model, and $\alpha_{Y_1}^2 = \alpha_{Y_1}^3 = 0, 0.1, 0.5$ and $\alpha_{Y_2}^3 = 0, 0.05, 0.05$ in the product-normal model. In the main set of simulations, all three markers were identical with regards to the aspects being studied. In additional simulations, we varied these aspects across the markers (details in Section 5 of supplementary material available at *Biostatistics* online) and also assessed the impact of increasing sample size to $n = 400$.

4.3. *Analysis of each generated data set and performance indicators*

In each scenario, 1000 data sets were generated. The target parameters, $\gamma_{Y_1}, \gamma_{Y_2}, \gamma_{Y_3}, \gamma_{Z_1}$, and γ_{Z_2}, were estimated from each data set by fitting the Cox or additive model with various approaches: using the true values of the markers at the event times according to the simulation model, which shows what one would obtain if perfect data were available ("True" analysis); the LOCF approach; the simple2S approach in the main set of simulations with low and high correlation; and the two MI strategies described, MIJM and unMIJM, performing $M = 5$ imputations and $G = 5$ iterations. In selected scenarios, we assessed the impact of excluding the interaction term $\widehat{\Lambda}(\tilde{T}_i) \times \boldsymbol{Z}_i$ from the imputation model, and of increasing the number of imputations to $M = 20$. For the Cox model only, in selected scenarios, we considered the JMbayes approach.

In each scenario, for each parameter $\gamma_.$, each estimator $\widehat{\gamma}_.$ and its standard error estimator $\widehat{\mathrm{SE}}(\widehat{\gamma}_.)$, we calculated the following indicators, where all means are taken across the 1000 generated data sets: the mean bias (MB) given by MB= mean$(\widehat{\gamma}_.) - \gamma_.$; the mean relative bias (MRB), given by MRB= $100 \times \{\mathrm{mean}(\widehat{\gamma}_.)/\gamma_. - 1\}$; the observed (empirical) standard error (ObsSE), given by the standard deviation of $\widehat{\gamma}_.$ across the 1000 data sets; the standardized bias (StdBias), given by StdBias=$100 \times [\{\mathrm{mean}(\widehat{\gamma}_.) - \gamma_.\}]/\mathrm{ObsSE}$; the mean estimated standard error (EstSE), EstSE= mean$\{\widehat{\mathrm{SE}}(\widehat{\gamma}_.)\}$; and the coverage probability (CP), given by the percentage of the 1000 data sets for which the 95% confidence interval (CI) contained the true value $\gamma_.$.

4.4. *Results*

Figures 1 and 2 show the standardized bias for γ_{Y_1} with five approaches across the main set of scenarios, for the additive hazard model and $K = 3$ markers generated from a correlated random effects model with low and high correlation, respectively. Table 1 shows detailed results for the additive model and both multiple marker models in selected scenarios. Next we summarize the results from the main set of simulations regarding estimation of $\gamma_{Y_1}, \gamma_{Y_2}$, and γ_{Y_3} with $n = 200$ and in the settings with low and high correlation for the additive model (a total of 108 scenarios), and further below for the Cox model. In the Supplementary material available at *Biostatistics* online we provide results for the case $K = 1$ (Section 3), further results for the main simulations (Section 4) and results for the additional simulations (Section 5), which led to similar conclusions.

With perfect data ("True" analysis), estimates obtained across all scenarios were approximately unbiased, standard errors were correctly estimated and CIs achieved nominal coverage (MRB between -23% and 12% and CP between 93% and 98% across all 108 scenarios; see also Table 1).

The LOCF approach led to negative bias in scenarios with measurement error or missingness, and the magnitude of the bias increased as the effect size increased and as these issues were more pronounced. The bias of LOCF was worst in scenarios with low correlation between the markers (MRB between -96%
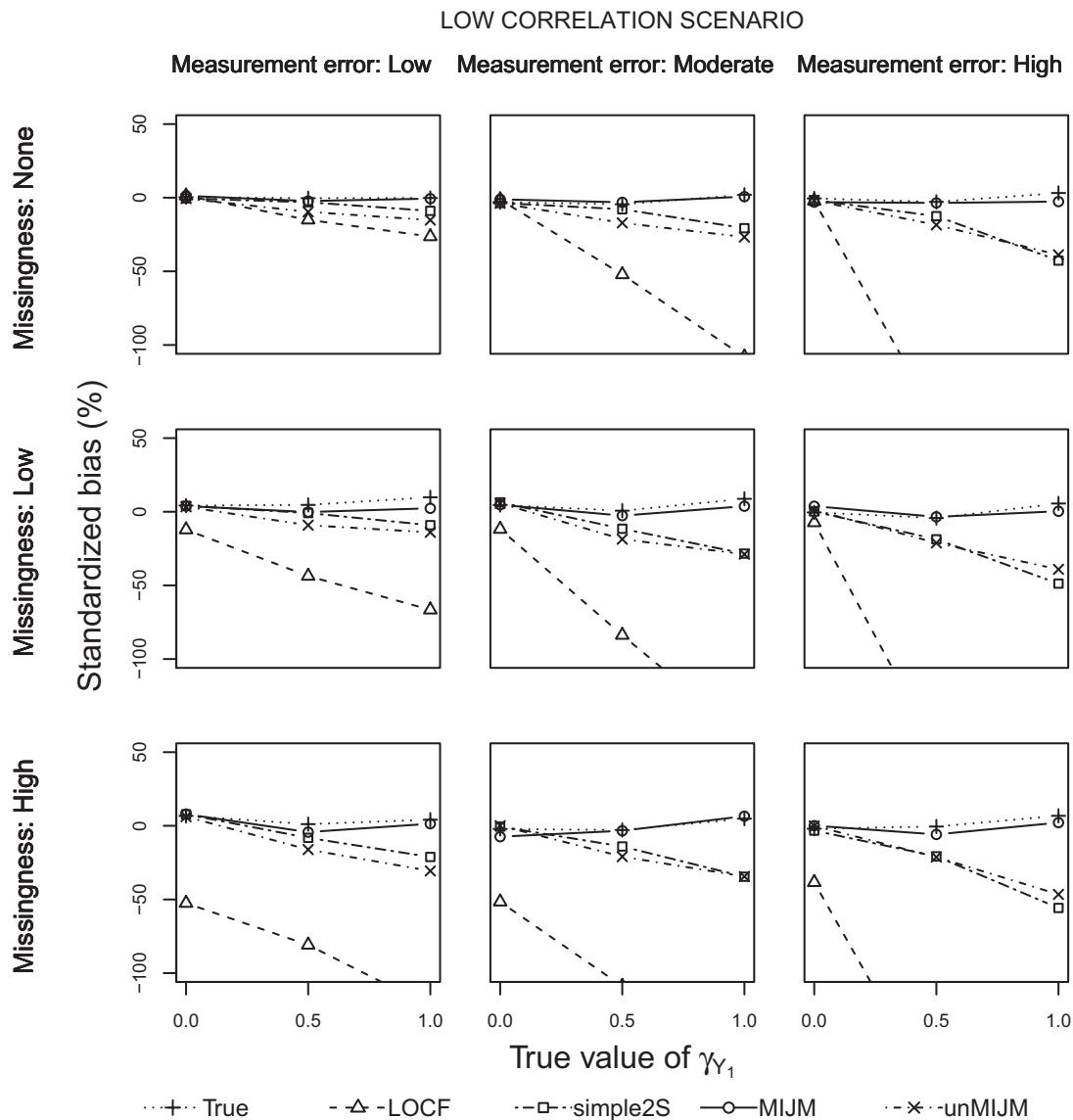
Fig. 1. Simulation study results: Standardized bias for the regression coefficient of marker 1, γ_{Y_1}, expressed in effect per 1000 units, in the additive model with low-correlated multiple markers generated from a correlated random effects model.

and -10% across all scenarios with low correlation compared with -89% to 2% with high correlation).
Notably, with missing measurements there was a negative bias even in the scenarios with null true effects
($\gamma_{Y_k} = 0$). The variance of the LOCF estimator decreased as the missingness and the measurement error
increased, failing to reflect the increased uncertainty in these scenarios, and being even lower than what was
achieved with perfect data. Together, the large biases and small standard errors translated into standardized
biases of very large magnitude (Figures 1 and 2). Estimated standard errors were close to their empirical
counterparts. Hence, solely as a result of the behavior of the regression coefficient estimator, this analysis
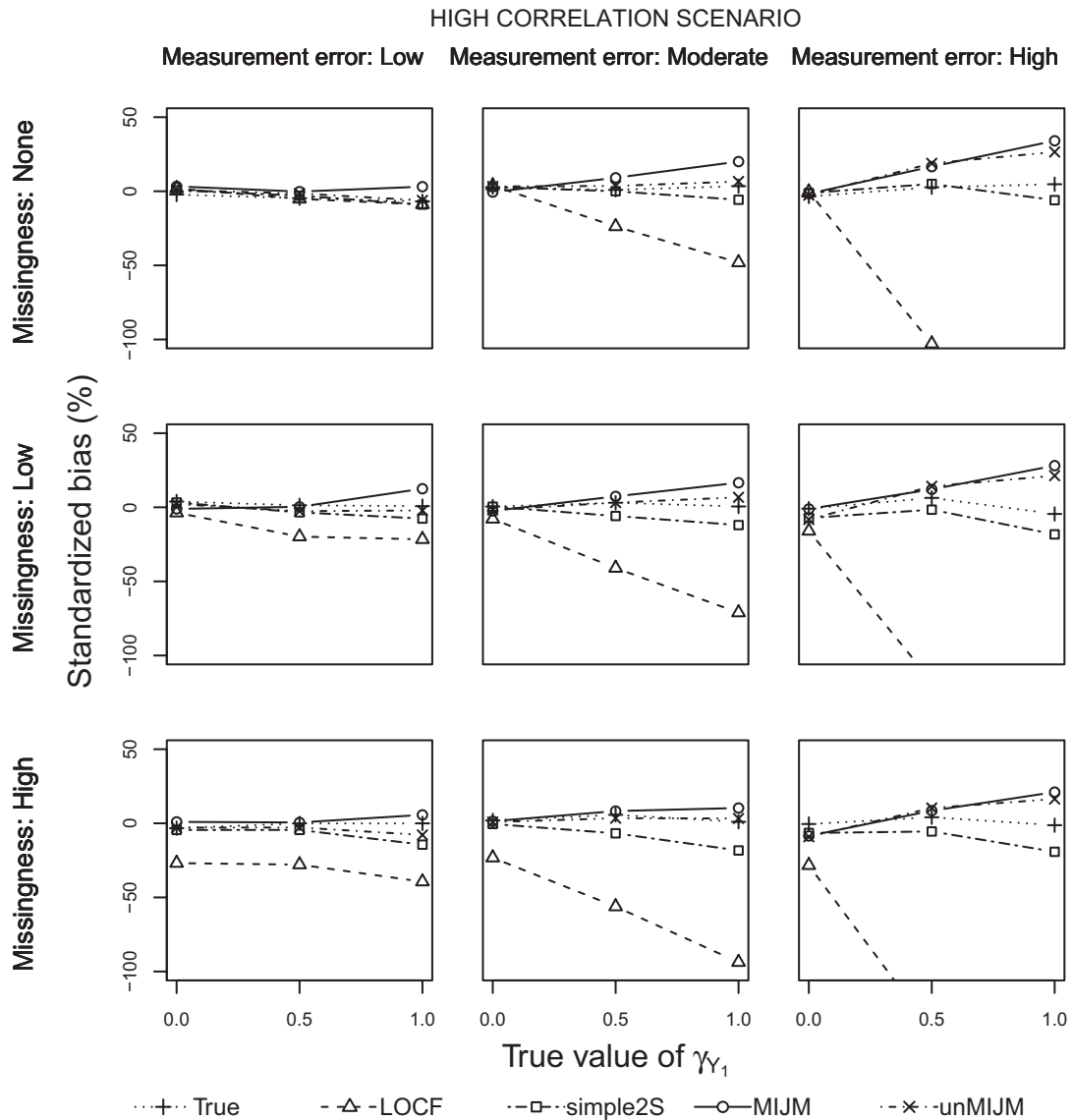yielded very poor coverage probability across most scenarios (CP as low as 3% in some scenarios).

Fig. 2. Simulation study results: Standardized bias for the regression coefficient of marker 1, γ_{Y_1}, expressed in effect per 1000 units, in the additive model with highly correlated multiple markers generated from a correlated random effects model.

The simple2S approach had reduced bias compared with LOCF, but exhibited negative bias in scenarios with low correlation, which increased as missingness and measurement error increased (MRB between -37% and 0% across these scenarios). The approach was approximately unbiased in scenarios with high correlation (MRB between -24% and 9%). Standard errors were generally higher than for LOCF but close to those with perfect data, thus also failing to reflect the increased uncertainty due to unobserved data. Standardized biases were lower than for LOCF but still important in settings with low correlation or high missingness (Figures 1 and 2). The standard error estimator was approximately unbiased (between -12%

and 15% difference relative to empirical standard errors). CPs were poor in settings with bias, ranging between 87% and 98% in low correlation scenarios and between 90% and 98% with high correlation.

The MIJM approach yielded approximately unbiased estimates across scenarios with low correlation (MRB between -27% and 16%). A positive bias was observed in some settings with high measurement error or strong effect size across scenarios with high correlation (MRB between -13% and 80%). These biases were less pronounced for the product-normal model (MRB between -13% and 59%) than the correlated random effects model (MRB between -3% and 80%), and biases in the product-normal model were highest for γ_{Y_1} possibly because Y^2 and Y^3 were generated with conditional models that more closely resemble the imputation model (maximal MRB 59% for γ_{Y_1} compared with 30% and 24% for γ_{Y_2} and γ_{Y_3}, respectively). Still, the MIJM estimator had the highest variance, which was always greater than that with perfect data and appropriately increased as missingness and measurement error increased. The variance was highest when the bias was highest, which resulted in consistently small standardized biases. Rubin's variance formula performed well in this analysis: although it led to slightly overestimated standard errors in some cases (up to 30% higher than empirical standard error), the CIs achieved coverages close to or somewhat higher than the nominal value across all scenarios (CP between 94% and 99% across the 108 scenarios). Notably, CPs remained satisfactory even in the cases of largest bias.

The unMIJM approach resulted in negative bias in scenarios with low correlation, which increased as missingness and measurement error increased (MRB between -38% and -6%). The estimator was approximately unbiased or had moderate bias across scenarios with high correlation (MRB between -20% and 31%). As for MIJM, in scenarios with high correlation, the bias of unMIJM was lower with the product-normal model (MRB between -20% and 15%) than with the correlated random effects model (MRB between -15% and 31%). The standard errors of the unMIJM estimator were close to those of the True and simple2S analyses. Standardized biases were thus generally larger in magnitude or similar to those of MIJM. For this analysis, Rubin's formula overestimated the standard error in most cases (between 10% and 73% higher than empirical standard errors), a consequence of the substantive-model-incompatibility implicit in this approach. CIs were thus over-conservative, ranging from 96% to 100% across all scenarios.

Results for the Cox model exhibited similar patterns. In the selected scenarios where it was considered, the JMbayes approach was approximately unbiased, standard errors were appropriately higher than for the True analysis, and standardized biases were small (see Table S3 of supplementary material available at *Biostatistics* online). This was to some extent expected for the correlated random effects model, which is the marker model assumed by JMbayes, but it is notable that this was also true for the product-normal model. On the other hand, there was a consistent underestimation of standard errors with this approach, particularly with the product-normal model. This may be due to misspecification by JMbayes of the marker model or B-splines providing a poor approximation of the baseline hazard. The biased standard error estimates led to poor CPs in some scenarios.

5. EXAMPLE: FRAMINGHAM HEART STUDY

We focused on assessing the associations between BMI, SBP and blood glucose levels, and all-cause mortality. Following Abdullah *and others* (2011), we fitted survival models for the time-to-death since recruitment including these time-varying markers as explanatory variables, in addition to some time-fixed baseline covariates (age, sex, country of birth, education level, smoking, alcohol use, and marital status). Analyses were restricted to the $n = 4881$ participants who did not have diabetes or cardiovascular disease at the baseline visit, and who had all three markers non-missing on at least one of their 28 two-yearly visits. The maximum follow-up time was 59.2 years, and 4552 participants had died (93%) with the remaining being censored on December 31, 2007 (see Section 6 of supplementary material available at *Biostatistics* online, for Kaplan–Meier curve and statistics on baseline factors).

Table 1. Simulation study results: Performance of estimators for additive model parameters in the scenario with moderate measurement error, low proportion of missing at-risk–visit measurements and strong effects of the markers on the hazard. Results for low and high correlations and each multiple marker model are shown. Effect sizes and standard errors are expressed per 1000 units

Analysis	Correlation	$\gamma_{\gamma_1} = 1$					$\gamma_{\gamma_2} = 1$					$\gamma_{\gamma_3} = 1$				
		MRB	StdBias	ObsSE	EstSE	CP	MRB	StdBias	ObsSE	EstSE	CP	MRB	StdBias	ObsSE	EstSE	CP
Correlated random effects model																
True	Low	4.3	8.8	0.5	0.5	96.0	3.1	6.2	0.5	0.5	94.2	−0.0	−0.0	0.5	0.5	94.7
LOCF		−54.8	−152.6	0.4	0.4	71.0	−55.4	−149.7	0.4	0.4	68.9	−57.3	−150.8	0.4	0.4	65.0
simple2S		−14.2	−28.4	0.5	0.5	94.6	−15.2	−29.4	0.5	0.5	93.1	−16.8	−31.5	0.5	0.5	92.6
MIJM		3.2	3.8	0.9	1.0	97.4	3.0	3.5	0.9	1.0	97.3	−1.5	−1.7	0.9	1.0	95.3
unMIJM		−15.0	−28.8	0.5	0.7	98.9	−17.0	−31.8	0.5	0.7	98.9	−18.1	−32.9	0.5	0.7	97.9
True	High	0.4	0.5	0.8	0.8	94.1	0.3	0.4	0.8	0.8	94.4	1.1	1.4	0.8	0.8	95.6
LOCF		−29.3	−71.1	0.4	0.4	88.5	−26.9	−65.5	0.4	0.4	88.4	−29.8	−74.2	0.4	0.4	89.6
simple2S		−9.0	−12.0	0.8	0.7	95.1	−9.1	−12.1	0.8	0.7	94.8	−12.2	−16.7	0.7	0.7	95.6
MIJM		27.8	16.6	1.7	1.8	96.5	29.2	17.6	1.7	1.8	95.6	25.3	15.6	1.6	1.7	96.7
unMIJM		5.5	6.8	0.8	1.1	98.9	4.2	5.3	0.8	1.1	98.4	2.1	2.6	0.8	1.1	99.2
Product-normal model																
True	Low	3.3	6.6	0.5	0.5	95.3	2.8	5.5	0.5	0.5	94.0	0.4	0.8	0.5	0.5	94.7
LOCF		−56.3	−149.9	0.4	0.4	66.9	−56.4	−149.1	0.4	0.4	66.8	−57.1	−151.5	0.4	0.4	66.7
simple2S		−14.4	−27.7	0.5	0.5	94.2	−16.1	−30.5	0.5	0.5	93.6	−17.1	−32.4	0.5	0.5	93.3
MIJM		1.3	1.5	0.9	1.0	96.4	−0.2	−0.2	0.9	1.0	96.6	0.5	0.5	0.9	1.0	96.4
unMIJM		−15.5	−28.6	0.5	0.7	98.3	−16.5	−30.1	0.5	0.7	98.0	−19.8	−36.5	0.5	0.7	98.4
True	High	7.5	12.9	0.6	0.6	95.2	−3.3	−6.7	0.5	0.5	94.8	−0.6	−1.3	0.5	0.5	95.1
LOCF		−30.0	−81.5	0.4	0.4	89.4	−42.3	−121.8	0.3	0.4	77.4	−38.8	−114.5	0.3	0.4	81.2
simple2S		−11.1	−19.3	0.6	0.6	95.8	−13.4	−27.8	0.5	0.5	94.4	−8.7	−18.2	0.5	0.5	95.1
MIJM		22.2	19.7	1.1	1.2	95.5	3.7	4.4	0.8	0.9	96.8	9.6	11.5	0.8	0.9	95.9
unMIJM		1.1	1.8	0.6	0.8	99.1	−14.6	−29.8	0.5	0.7	98.3	−9.5	−19.8	0.5	0.7	99.2

MRB, mean relative bias (%); StdBias, standardized bias (%); ObsSE, observed (empirical) standard error; EstSE, mean estimated standard error; CP, coverage probability (%).
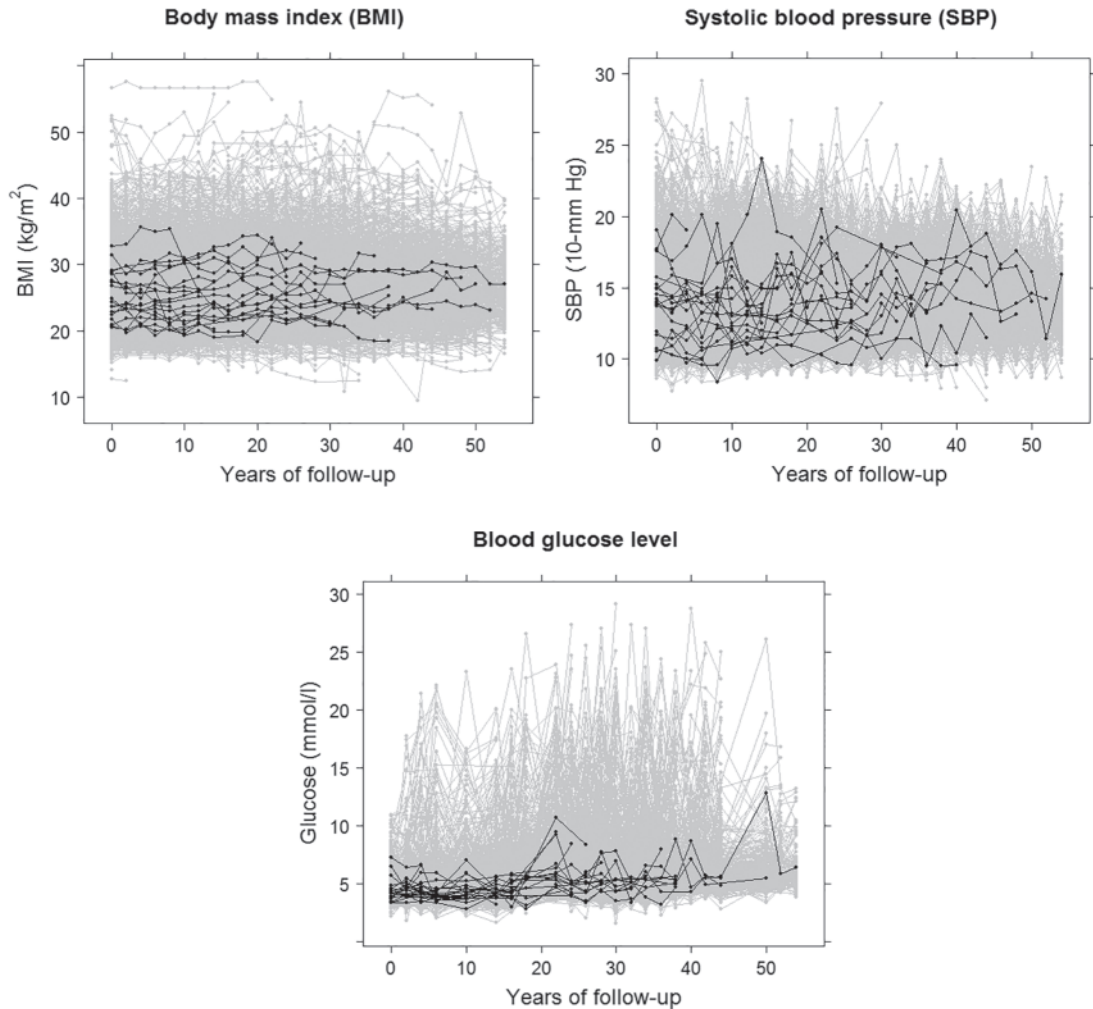
Fig. 3. Trajectories based on available measurements for three time-dependent markers in the Framingham Heart Study, for all individuals (gray) and for a random sample of ten individuals (black).

Figure 3 shows the profiles of the three time-varying markers for a random subset of 10 participants. These markers span the three scenarios of measurement error explored in our simulation study, with BMI having a low degree of error (ICC $= 0.89$ as estimated from the available data), SBP having a moderate degree of error (ICC $= 0.63$) and blood glucose having a high degree of error (ICC $= 0.15$). The proportion of missing at-risk-visit measurements was low for BMI and SBP (21% and 24%, respectively) and high for blood glucose (41%). These characteristics suggest the need for approaches beyond LOCF. Marginal correlations estimated from the observed data were: corr(BMI,SBP) $= 0.24$, corr(BMI,glucose) $= 0.13$ and corr(SBP,glucose) $= 0.14$.

We fitted Cox and additive hazard models as in (2.1) and (2.2), respectively, using LOCF, simple2S, MIJM, unMIJM, and, in the case of the Cox model, JMbayes. In the latter four approaches, time-trends for each marker were modeled using cubic polynomials of time (which provided superior or similar fit compared with natural cubic splines), and the models included random intercepts and slopes. Based on

Table 2. *Application to the Framingham Heart Study: Estimated mortality hazard ratios (HR) and differ-ences (HD) for an increase in one unit of each time-dependent marker, obtained by fitting multivariable Cox and additive hazard models, respectively, using various approaches to incorporate these imperfectly observed time-dependent covariates*

Marker	Analysis	Cox model			Additive model		
		HR	95% CI	*p*-value	HD[a]	95% CI	*p*-value
BMI	LOCF	0.97	(0.97; 0.98)	<0.001	−113	(−135; −91)	<0.001
(kg/m²)	simple2S	0.98	(0.98; 0.99)	<0.001	−110	(−134; −86)	<0.001
	MIJM	0.95	(0.93; 0.96)	<0.001	−200	(−233; −166)	<0.001
	unMIJM	0.98	(0.97; 0.99)	<0.001	−106	(−133; −80)	<0.001
	JMbayes[b]	0.98	(0.97; 0.99)	<0.001	—	—	—
SBP	LOCF	1.02	(1.01; 1.04)	<0.001	2	(−47; 50)	0.95
(10 mmHg)	simple2S	1.15	(1.13; 1.17)	<0.001	284	(220; 348)	<0.001
	MIJM	0.97	(0.94; 1.02)	0.21	−186	(−316; −56)	0.01
	unMIJM	1.13	(1.10; 1.16)	<0.001	278	(195; 361)	<0.001
	JMbayes[b]	1.14	(1.12; 1.16)	<0.001	—	—	—
Glucose	LOCF	1.05	(1.04; 1.07)	<0.001	262	(181; 342)	<0.001
(mmol/L)	simple2S	1.13	(1.11; 1.15)	<0.001	719	(584; 854)	<0.001
	MIJM	1.18	(1.14; 1.22)	<0.001	971	(721; 1221)	<0.001
	unMIJM	1.11	(1.08; 1.13)	<0.001	555	(421; 689)	<0.001
	JMbayes[b]	1.14	(1.12; 1.15)	<0.001	—	–	—

BMI, body mass index; SBP, systolic blood pressure; CI, confidence interval;
HR, hazard ratio; HD, hazard difference.
[a]Units: events per 100 000 person-years.
[b]This approach was applicable only to the Cox model.

relative efficiency calculations and graphical assessments, respectively, $M = 20$ imputations and $G = 5$ iterations were used for MIJM and unMIJM. Indeed, results with $M = 50$ were essentially the same (not shown).

Table 2 shows the results. For BMI, all approaches suggest evidence of a protective association with each unit increase, independent of current SBP and blood glucose levels, although MIJM suggests a stronger association both in absolute and relative terms. For SBP, in relative terms all approaches suggest evidence of a weak harmful independent association with increasing SBP except MIJM, which does not show evidence of an association. In absolute terms, results are discrepant across the approaches: LOCF does not show evidence of an association, MIJM indicates a weak protective association with increasing SBP and simple2S and unMIJM display the opposite. For blood glucose, all approaches suggest evidence of a harmful independent association with increasing glucose levels in both relative and absolute terms, although MIJM suggests a much stronger relation. The results of this illustrative example are to be interpreted with caution since the fit of these simple Cox and additive hazard models could perhaps be improved by modeling U-shaped associations, particularly for SBP (see Section 6 of supplementary material available at *Biostatistics* online).

6. CONCLUDING REMARKS

We developed a refined two-stage joint modeling approach, MIJM, to incorporate multiple markers in two hazard rate models. Simulation results showed that MIJM performed well with one or multiple markers,

with small standardized biases and CIs that achieved nominal coverage across all scenarios studied and for all multiple-marker/survival model combinations considered. The latter shows that the absolute bias of MIJM in some extreme scenarios is of little practical significance due to the accompanying increased variability (Collins *and others*, 2001).

The unMIJM approach, which differed from MIJM only in the way the time-to-event outcome was included in the imputation model, had a performance that was variable across scenarios and overall poorer than that of MIJM. This difference emphasizes the importance of appropriately addressing issue (iii) when using two-stage approaches. Further, the implicit less-refined approximation of unMIJM yielded overly-conservative CIs in most scenarios due to systematic bias in the MI variance estimator. The unMIJM approach may, however, have value in settings with highly correlated markers and low measurement error. The simple2S approach also showed variable performance across scenarios, although it performed well in scenarios with high correlation and low missingness.

Full joint modeling performed well in terms of bias, but standard error estimates seemed vulnerable to misspecification of the full joint distribution, which affected CPs. More generally, this approach is computationally challenging: at the time of writing we were unable to identify any readily available full joint modeling software, Bayesian or frequentist, that accommodated multiple markers and both semi-parametric hazard models considered in this paper. While it was possible to apply the approach assuming a proportional hazards model with spline-modeled baseline, we could only do this in selected scenarios due to extended running time. All approaches seemed clearly preferable to LOCF.

The MIJM approach can be easily implemented for any number of markers using mainstream software, and the R package `survtd` reduces this task to a single function call. Of note, the package can in addition be used to apply unMIJM, simple2S, and LOCF as easily. The MIJM approach is also computationally fast, demonstrating substantially lower running times than JMbayes.

Future work could include refinements of MIJM to better accommodate strongly correlated markers or to tackle other time-to-event models. Assessment of the MIJM approach in the context of unplanned rather than planned visiting times would also be of interest. While there are no practical obstacles to applying the approach in that setting, assumptions about ignorability of the visiting process would be required, and the modified event indicator would be defined in terms of the last available measurement rather than the last planned visit before end of follow-up. Thus, the accuracy of the approximation underlying the approach could be affected, e.g. if there were high variability in the delays between visiting times.

To conclude, the proposed MIJM approach provides a pragmatic yet principled solution to a common problem in health research. Some inevitable approximations are involved in the imputation modeling, but the approach was robust in our simulations, which explored a wide range of scenarios using structures and parameter settings that mimicked many aspects of the FHS and could be expected to reflect other similar studies.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

References

ABDULLAH, A., WOLFE, R., STOELWINDER, J. U., DE COURTEN, M., STEVENSON, C., WALLS, H. L. AND PEETERS, A. (2011). The number of years lived with obesity and the risk of all-cause and cause-specific mortality. *International Journal of Epidemiology* **40**, 985–996.

ALBERT, P. S. AND SHIH, J. H. (2010). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics* **66**, 983–987.

ANDERSEN, P. K. AND LIESTØL, K. (2003). Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics* **4**, 633–649.

BATES, D, MÄCHLER, M, BOLKER, B. M. AND WALKER, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48.

BYCOTT, P. AND TAYLOR, J. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statistics in Medicine* **17**, 2061–2077.

CHI, Y. Y. AND IBRAHIM, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* **62**, 432–445.

COLLINS, L. M., SCHAFER, J. L. AND KAM, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* **6**, 330–351.

COOPER, N. J., LAMBERT, P. C., ABRAMS, K. R. AND SUTTON, A. J. (2007). Predicting costs over time using Bayesian Markov chain Monte Carlo methods: an application to early inflammatory polyarthritis. *Health Economics* **16**, 97–56.

DANIELS, M. J. AND HOGAN, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton, FL: CRC Press.

DE GRUTTOLA, VI AND TU, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.

FAUCETT, C. L. AND THOMAS, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* **15**, 1663–1685.

FIEUWS, S. AND VERBEKE, G. (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statistics in Medicine* **23**, 3093–3104.

GOULD, A. L., BOYE, M. E., CROWTHER, M. J., IBRAHIM, J. G., QUARTEY, G., MICALLEF, S. AND BOIS, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA bayesian joint modeling working group. *Statistics in Medicine* **34**, 2181–2195.

HANSON, T. E., BRANSCUM, A. J. AND JOHNSON, W. O. (2011). Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. *Lifetime Data Analysis* **17**, 3–28.

HICKEY, G. L., PHILIPSON, P., JORGENSEN, A. AND KOLAMUNNAGE-DONA, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology* **16**, 117.

LIN, D. Y. AND YING, Z. L. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

MORENO-BETANCUR, M. (2017). *survtd: Survival Analysis with Time-Dependent Covariates.* https://github.com/moreno-betancur/survtd.

MORENO-BETANCUR, M. AND CHAVANCE, M. (2016). Sensitivity analysis of incomplete longitudinal data departing from the missing at random assumption: methodology and application in a clinical trial with drop-outs. *Statistical Methods in Medical Research* **25**, 1471–1489.

RIZOPOULOS, D. (2016a). *Multivariate Joint Models for Multiple Longitudinal Outcomes and a Time-to-Event.* https://www.r-bloggers.com/multivariate-joint-models-for-multiple-longitudinal-outcomes-and-a-time-to-event/.

RIZOPOULOS, D. (2016b). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software* **72**, 1–45.

RIZOPOULOS, D. AND GHOSH, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine* **30**, 1366–1380.

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

SEAMAN, S., GALATI, J., JACKSON, D. AND CARLIN, J. B. (2013). What is meant by missing at random? *Statistical Science* **28**, 257–268.

SPIEGELHALTER, D. J. (1998). Bayesian graphical modelling: a case-study in monitoring health outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **47**, 115–133.

THERNEAU, T. AND GRAMBSCH, P. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

VAN BUUREN, S, BOSHUIZEN, H AND KNOOK, D. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistic in Medicine* **18**, 681–694.

VAN BUUREN, S. AND GROOTHUIS-OUDSHOORN, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software* **45**, 1–67.

WHITE, I. R. AND ROYSTON, P. (2009). Imputing missing covariate values for the Cox model. *Statistics in Medicine* **28**, 1982–1998.

WU, L., LIU, W., YI, G. Y. AND HUANG, Y. (2012). Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics* **2012**, 1–17.

WULFSOHN, M. S. AND TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

XU, J. AND ZEGER, S. L. (2001). The evaluation of multiple surrogate endpoints. *Biometrics* **57**, 81–87.

YE, W., LIN, X. and TAYLOR, J. M. G. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data–a two-stage regression calibration approach. *Biometrics* **64**, 1238–46.