Statistical modelling of competing risks with incomplete data: with applications to allogeneic stem cell transplantation

Bonneville, E.F.

Chapter 5

Multiple imputation of missing covariates when using the Fine–Gray model

Chapter based on: **Bonneville, E. F.**, Beyersmann, J., Keogh, R. H., et al. (2024) Multiple imputation of missing covariates when using the Fine–Gray model. arXiv:2405.16602. arXiv. DOI: 10.48550/arXiv.2405.16602.

Abstract

The Fine–Gray model for the subdistribution hazard is commonly used for estimating associations between covariates and competing risks outcomes. When there are missing values in the covariates included in a given model, researchers may wish to multiply impute them. Assuming interest lies in estimating the risk of only one of the competing events, this paper develops a substantive-model-compatible multiple imputation approach that exploits the parallels between the Fine–Gray model and the standard (single-event) Cox model. In the presence of right-censoring, this involves first imputing the potential censoring times for those failing from competing events, and thereafter imputing the missing covariates by leveraging methodology previously developed for the Cox model in the setting without competing risks. In a simulation study, we compared the proposed approach to alternative methods, such as imputing compatibly with cause-specific Cox models. The proposed method performed well (in terms of estimation of both subdistribution log hazard ratios and cumulative incidences) when data were generated assuming proportional subdistribution hazards, and performed satisfactorily when this assumption was not satisfied. The gain in efficiency compared to a complete-case analysis was demonstrated in both the simulation study and in an applied data example on competing outcomes following an allogeneic stem cell transplantation. For individual-specific cumulative incidence estimation, assuming proportionality on the correct scale at the analysis phase appears to be more important than correctly specifying the imputation procedure used to impute the missing covariates.

5.1 Introduction

The presence of missing covariate data continues to be pervasive across biomedical research. Among the many existing approaches for dealing with missing covariate data, multiple imputation (MI) methods in particular have become increasingly popular in practice (Carpenter and Smuk, 2021). Compared to a complete-case analysis (CCA), MI can provide inferences that are both less biased and more efficient, under certain missingness mechanisms and given that the imputation models are appropriately specified (Sterne *et al.*, 2009).

The most common approach to MI is to specify and fit univariate regression models for partially observed covariates, from which imputations are then generated. Ideally, each one of these imputation models should be compatible with the substantive model of interest. That is, the assumptions made by both models should not conflict with each other, e.g. the imputation model should at least feature the remaining substantive model covariates, as well as the outcome. We refer to an imputation model as being 'directly specified' when substantive model covariates and outcome variable(s), or any transformations thereof, are included explicitly as predictors in the imputation model. In settings with missingness spanning multiple covariates, the specification of a joint distribution via a set of directly specified imputation models is more commonly known as MICE (multivariate imputation by chained equations, van Buuren *et al.*, 1999).

In the context of cause-specific Cox proportional hazards models (Prentice *et al.*, 1978), it has been shown that the imputation model for a partially observed covariate should at least include as predictors the other covariates from the substantive model, together with the cause-specific cumulative hazard and event indicator for each competing risk (Bonneville *et al.*, 2022). Analogously to the standard single-event survival setting, this directly specified imputation model is generally only approximately compatible with the proportional hazards substantive model (White and Royston, 2009). Concretely, when the outcome model assumes proportional hazards, the conditional distribution of a partially observed covariate modelled using MICE is only an approximation of the 'true' (i.e. implied assuming the substantive model is correctly specified) conditional distribution of the partially observed covariate given the outcome and other substantive model covariates. If imputed values can instead be directly sampled from the latter distribution, it would ensure compatibility between analysis and imputation model. This alternative 'indirect' way of obtaining imputations is referred to as the substantive-model-compatible imputation (SMC-FCS, Bartlett *et al.*, 2015) approach, and it was extended by Bartlett and Taylor (2016) to accommodate cause-specific Cox substantive models. In terms of estimating cause-specific hazard ratios, simulation studies have shown that the SMC-FCS approach tends to outperform MICE in cases when the substantive model is correctly specified (Bartlett and Taylor, 2016; Bonneville *et al.*, 2022).

When a Fine–Gray subdistribution hazard model (Fine and Gray, 1999) is the substantive model of interest, there has to our knowledge been no research on how one should specify an imputation model for a missing covariate (Lau and Lesko, 2018). Nevertheless, MICE is still being used in the presence of missing covariates when the substantive model is a Fine–Gray model, particularly in the context of prediction models. While the structure of the imputation model is rarely reported, articles which do describe their imputation procedure appear to use different approaches. For example, in the prognostic Fine–Gray model presented by Archer *et al.* (2022) (where the primary outcome was time to serious fall resulting in hospital admission or death, with competing death due to other causes), the imputation model for a missing covariate contained the other substantive model covariates, and the cause-specific cumulative hazard and event indicator for each competing risk. In contrast, the MICE procedure reported as part of the prognostic models presented by Clift *et al.* (2020) used cumulative subdistribution hazards in the imputation model. Heuristically, it would seem the latter approach is more consistent with the substantive model, as the former imputes approximately compatibly with a cause-specific Cox model structure rather than the Fine–Gray model structure.

In this work, we extend the SMC-FCS approach for missing covariates to accommodate a Fine–Gray substantive model for one of the competing events. In the presence of independent and identically distributed censoring times that are stochastically independent of the competing risks process (i.e. the 'random censoring' assumption, Beyersmann *et al.*, 2012), the core idea is to multiply impute the potential censoring times for individuals failing from competing events in a first step (Ruan and Gray, 2008), and thereafter use existing SMC-FCS methodology (originally developed for the standard Cox model, Bartlett *et al.*, 2015) to impute the missing covariates in a second step.

The structure of the manuscript is as follows. We introduce competing risks notation in Section 5.2. In Section 5.3 we outline the proposed method, and thereafter assess its performance in a simulation study in Section 5.4. We also provide an illustrative analysis using a dataset from the field of allogeneic hematopoietic stem cell transplantation (alloHCT) in Section 5.5. Finally, findings are discussed in Section 5.6, together with recommendations on how to impute covariates in competing risks settings more generally.

5.2 Notation

We consider a setting in which individuals experience only one of K competing events. We denote the event time as $\tilde{T}$, and the competing event indicator as $\tilde{D} \in \{1, ..., K\}$. In practice, individuals are subject to some right-censoring time C, meaning we only observe realisations (t_i, d_i) of $T = \min(C, \tilde{T})$ and $D = I(\tilde{T} \leq C)\tilde{D}$, where $I(\cdot)$

is the indicator function and $D = 0$ indicates a right-censored observation. The cause-specific hazard for the k^{th} event is defined as

$$h_k(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq \tilde{T} < t + \Delta t, \tilde{D} = k \mid \tilde{T} \geq t)}{\Delta t}.$$

These hazards together make up the event-free survival function,

$$P(\tilde{T} > t) = \exp\left\{ -\sum_{k=1}^{K} \int_0^t h_k(u)\,\mathrm{d}u \right\} = \exp\left\{ -\sum_{k=1}^{K} H_k(t) \right\},$$

assuming the distribution of T is continuous, and $H_k(t)$ is the cause-specific cumulative hazard function for the k^{th} event. The cause-specific cumulative incidence function is then defined as

$$F_k(t) = P(\tilde{T} \leq t, \tilde{D} = k) = \int_0^t h_k(u) S(u-)\,\mathrm{d}u,$$

where $S(u-)$ is the event-free survival probability just prior to u.

The subdistribution hazard for the k^{th} event is defined as

$$\begin{aligned} \lambda_k(t) &= \frac{-\,\mathrm{d}\log\{1 - F_k(t)\}}{\mathrm{d}t}, \\ &= \frac{\mathrm{d}F_k(t)}{\mathrm{d}t} \times \{1 - F_k(t)\}^{-1}, \end{aligned}$$

which can be thought of as the hazard for the improper random variable $\tilde{V}_k = I(\tilde{D} = k) \times \tilde{T} + I(\tilde{D} \neq k) \times \infty$, for which we can write $F_k(t) = P(\tilde{V}_k \leq t)$ (Beyersmann *et al.*, 2012). The probability mass at infinity makes $\tilde{V}_k$ improper, i.e. that its density function does not integrate to one.

Suppose interest lies in modelling the cumulative incidence of one of the competing events, say $D = 1$, conditional on (time-fixed) covariates $\mathbf{Z}$. The Fine–Gray model for cause 1 can be written as

$$\lambda_1(t \mid \mathbf{Z}) = \lambda_{01}(t) \exp(\boldsymbol{\beta}^\top \mathbf{Z}),$$

with $\lambda_{01}(t)$ being the subdistribution baseline hazard function and $\boldsymbol{\beta}$ representing the effects of covariates $\mathbf{Z}$ on the subdistribution hazard. The cumulative incidence function for cause 1 can then be written as

$$F_1(t \mid \mathbf{Z}) = 1 - \exp\left\{ -\exp(\boldsymbol{\beta}^\top \mathbf{Z}) \int_0^t \lambda_{01}(u)\,\mathrm{d}u \right\},$$

where $\int_0^t \lambda_{01}(u)\,\mathrm{d}u = \Lambda_{01}(t)$ is the cumulative baseline subdistribution hazard. If we define a baseline cumulative incidence function $F_{01}(t) = 1 - \exp\{-\Lambda_{01}(t)\}$ (i.e. the cumulative incidence when $\mathbf{Z} = 0$), the model can also be written as

$$F_1(t \mid \mathbf{Z}) = 1 - \{1 - F_{01}(t)\}^{\exp(\boldsymbol{\beta}^\top \mathbf{Z})}. \tag{5.1}$$

In the presence of random right censoring, the Fine–Gray model is usually fitted by maximising a partial likelihood that uses time-dependent inverse probability of censoring weights (IPCW) (Fine and Gray, 1999).

5.3 MI approaches with a Fine–Gray substantive model

We consider a setting with p partially observed covariates $X = X_1, ..., X_p$, q fully observed covariates $Z = Z_1, ..., Z_q$, and $K = 2$ competing events. We assume that (possibly conditional on Z) censoring is independent of both X and the competing risks outcomes $\tilde{T}, \tilde{D}$. We furthermore let X^{obs} and X^{mis} respectively denote the observed and missing components of X for an individual, and let R be the vector of observation indicators (equal to 1 if the corresponding element of X is observed, or equal to 0 if it is missing).

The substantive model of interest is $\lambda_1(t \mid X, Z) = \lambda_{01}(t)\exp\{g(X, Z; \boldsymbol{\beta})\}$, which is a Fine–Gray model for cause 1, and where $g(X, Z; \boldsymbol{\beta})$ is a function of X and Z, parametrised by $\boldsymbol{\beta}$. In this section, we provide an overview of possible approaches for imputing each partially observed X_j. These imputation models can then be 'chained' together as described in Sections 4 and 5 of the work by Bartlett *et al.* (2015). In addition to an approach which imputes compatibly with the assumed substantive model, we also consider alternative methods which are either a) only approximately compatible with the substantive model; or b) impute assuming a different underlying competing risks structure (i.e. cause-specific proportional hazards). We require that the proposed approaches be valid under the missing-at-random (MAR) assumption, that is, $P(R \mid T, D, X, Z) = P(R \mid T, D, X^{\text{obs}}, Z)$.

5.3.1 MI based on cause-specific hazards models

5.3.1.1 CS-SMC

A first MI approach to consider is to impute compatibly with cause-specific Cox models, despite the substantive model of interest being a Fine–Gray model for cause 1. As

described by Bartlett and Taylor (2016), this method relies on the substantive-model-compatible imputation density for X_j, given by

$$f(X_j \mid T, D, X_{-j}, Z) \propto f(T, D \mid X, Z) f(X_j \mid X_{-j}, Z), \tag{5.2}$$

where X_{-j} refers to the components of X after removing X_j, and $f(\cdot)$ is a density function. For example, $f(T, D \mid X, Z)$ is used as shorthand notation for $f_{T,D \mid X,Z}(t, d \mid x, z)$, that is, the density function for the conditional distribution $T, D \mid X, Z$, evaluated at (t, d) for given values x and z.

In practice, the substantive model $f(T, D \mid X, Z; \psi)$ assumed for $f(T, D \mid X, Z)$ is a cause-specific Cox model (one for each competing risk). Therefore, ψ ($\psi \in \Psi$) contains the cumulative baseline hazards and log hazard ratios for each cause-specific hazard. A model $f(X_j \mid X_{-j}, Z; \phi)$ indexed by ϕ ($\phi \in \Phi$), is also assumed for $f(X_j \mid X_{-j}, Z)$. The idea is then to sample candidate imputed values for the missing X_j using $f(X_j \mid X_{-j}, Z; \phi)$, and accept these if they also represent draws from a density proportional to $f(T, D \mid X, Z; \psi) f(X_j \mid X_{-j}, Z; \phi)$. We refer to this method as the cause-specific SMC-FCS approach (CS-SMC).

5.3.1.2 CS-Approx

The approximately compatible analogue to the cause-specific SMC-FCS approach is described by Bonneville *et al.* (2022). As briefly described in the introduction, this approach involves directly specifying an imputation model $f(X_j \mid T, D, X_{-j}, Z; \alpha)$ for $f(X_j \mid T, D, X_{-j}, Z)$. In order to ensure approximate compatibility with assumed cause-specific Cox substantive models, the imputation model should include as predictors X_{-j}, Z, D (as a factor variable), and the (marginal, as obtained using the Nelson–Aalen estimator) cause-specific cumulative hazard for each cause $\hat{H}_k(T)$, evaluated at an individual's event or censoring time. We refer to this method as approximately compatible cause-specific MICE (CS-Approx).

5.3.1.3 MI based on the relation between the cause-specific and subdistribution hazards

The imputations generated by the CS-SMC and CS-Approx approaches will typically not be consistent with the assumption of proportional subdistribution hazards for cause 1 made by the substantive model of interest. This is because, for cause 1, proportionality on the cause-specific hazard scale will generally imply non-proportionality on the subdistribution hazard scale (Beyersmann *et al.*, 2012). One can derive the functional form of these time-varying covariate effects on the subdistribution hazard scale by using the relation between the subdistribution hazard and the cause-specific hazards (Putter *et al.*, 2020). The CS-SMC and CS-Approx approaches can therefore be thought

of as procedures to impute (approximately) compatibly with a Fine–Gray model with time-varying covariate effects, the functional form of which are determined by the assumptions made for the cause-specific Cox models of each competing event.

A relevant question at this point is whether the relation between cause-specific and subdistribution hazards can instead be used as part of a procedure to impute compatibly with proportional subdistribution hazards for cause 1. In order to motivate such a procedure, we first note that the conditional density of the observed outcome given covariates used in Equation 5.2 can be written both in terms of cause-specific hazards, and in terms of the cumulative incidence functions, as

$$\begin{aligned} f(T, D \mid X, Z) &= \{h_1(T \mid X, Z) S(T \mid X, Z)\}^{I(D=1)} \{h_2(T \mid X, Z) S(T \mid X, Z)\}^{I(D=2)} \\ &\quad \times S(T \mid X, Z)^{1-I(D=1)-I(D=2)}, \\ &= f_1(T \mid X, Z)^{I(D=1)} f_2(T \mid X, Z)^{I(D=2)} \\ &\quad \times \{1 - F_1(T \mid X, Z) - F_2(T \mid X, Z)\}^{1-I(D=1)-I(D=2)}, \end{aligned} \tag{5.3}$$

with $f_k(t \mid X, Z) = \mathrm{d}F_k(t \mid X, Z)/\,\mathrm{d}t$ known as the 'subdensity' for cause k (Gray, 1988). These subdensities, in turn, can be expressed in terms of the subdistribution hazard, as

$$\begin{aligned} f_k(t \mid X, Z) &= \lambda_k(t \mid X, Z)\{1 - F_k(t \mid X, Z)\}, \\ &= \lambda_k(t \mid X, Z) \exp\{-\Lambda_k(t \mid X, Z)\}, \end{aligned} \tag{5.4}$$

where $\Lambda_k(t \mid X, Z)$ is the cumulative subdistribution hazard for cause k conditional on X and Z. Specifying a Fine–Gray model for cause 1 is an assumption regarding only part of Equation 5.3, namely for any terms involving $f_1(T \mid X, Z)$. The practical implication of this is that Equation 5.2 cannot be used to impute the missing X_j without making assumptions about cause 2. One could for example assume (for imputation purposes) a cause-specific Cox model for cause 2, derive the implied $h_1(t \mid X, Z)$ using the relation between the subdistribution hazard and the cause-specific hazards, and then use both cause-specific hazards to evaluate $f(T, D \mid X, Z)$ in Equation 5.2.

Given that a Fine–Gray model is assumed for cause 1, some computational difficulties can be encountered while making assumptions for cause 2. For example, specifying a Fine–Gray model also for cause 2 in the imputation procedure could result in the total failure probability at an observed event time $F_1(T \mid X, Z) + F_2(T \mid X, Z)$ exceeding 1, meaning we would not be able to draw imputed values using Equation 5.2 for high-risk individuals (Austin *et al.*, 2021). An additional example concerns the approach described in the previous paragraph, where $h_1(t \mid X, Z)$ is derived based on $h_2(t \mid X, Z)$ and $\lambda_1(t \mid X, Z)$. The numerical integration step generally needed to compute $h_1(t \mid X, Z)$ could make the overall imputation procedure rather computationally inefficient. More details on potential issues when specifying a model for cause 2 when a Fine–Gray model is assumed for cause 1 can be found in Bonneville *et al.* (2024)

The above points mean that it is desirable to use an alternative approach which avoids having to specify a model for the cause-specific (or subdistribution) hazard of cause

2. In the next subsection, we propose a SMC-FCS approach assuming a Fine–Gray substantive model for cause 1, which avoids making explicit modelling assumptions concerning cause 2.

5.3.2 MI based on the Fine–Gray model

5.3.2.1 FG-SMC

Suppose for now that the potential censoring time C is known for all individuals. This is for example the case when there is a fixed end of study date (i.e. 'administrative' censoring), and no additional random right-censoring. Fine and Gray (1999) referred to these kind of data as 'censoring complete', since the subdistribution at-risk process is known. Equivalently, the 'observed' subdistribution random variable for cause 1 (henceforth referred to as 'subdistribution time'), $V = I(D = 1) \times T + I(D \neq 1) \times C$, is known for all individuals. In turn, this implies that (with complete covariate data), the Fine–Gray model can be estimated by fitting a standard Cox model with outcome V and event indicator $I(D = 1)$.

Consequently, an intuitive approach to imputing the missing X_j in our setting might therefore be to apply existing SMC-FCS methodology for standard Cox models (see section 6.3 of Bartlett *et al.*, 2015), but instead using V and $I(D = 1)$ as our outcome variables. We refer to this method as Fine–Gray SMC-FCS (FG-SMC). The substantive-model-compatible imputation density is now

$$f(X_j \mid V, D, Z) \propto f(V, D \mid X, Z) f(X_j \mid X_{-j}, Z), \tag{5.5}$$

where the conditional density of the observed outcome given the covariates can be written as

$$\begin{aligned} f(V, D \mid X, Z) &= f_1(V \mid X, Z)^{I(D=1)} \{1 - F_1(V \mid X, Z)\}^{I(D\neq 1)}, \\ &= [\lambda_1(V \mid X, Z) \exp\{-\Lambda_1(V \mid X, Z)\}]^{I(D=1)} \exp\{-\Lambda_1(V \mid X, Z)\}^{I(D=0)} \\ &\quad \exp\{-\Lambda_1(V \mid X, Z)\}^{I(D=2)}, \\ &= \lambda_1(V \mid X, Z)^{I(D=1)} \exp\{-\Lambda_1(V \mid X, Z)\}, \end{aligned} \tag{5.6}$$

using Equation 5.4 and the fact that $f_1(V \mid X, Z)^{I(D=1)} = f_1(T \mid X, Z)^{I(D=1)}$. Note that while Equation 5.5 and Equation 5.6 depend only on $I(D = 1)$, we still use D in the notation to make the contribution of those failing from cause 2 to the density explicit, which is relevant for the upcoming sections. Importantly, this procedure relies on a stronger MAR assumption (compared to the one introduced at the beginning of Section 5.3), namely $P\{R \mid V, I(D = 1), X, Z\} = P\{R \mid V, I(D = 1), X^{\text{obs}}, Z\}$. In essence, we ignore any terms involving $f_2(T \mid X, Z)$ in Equation 5.2 based on the assumption that missingness in X does not depend on either $I(D = 2)$ or the failure time for those failing from cause 2.

5.3.2.2 FG-Approx

The form of Equation 5.6 mirrors the likelihood in the standard Cox context, which can be obtained by replacing $\lambda_1(V \mid X, Z)$ with the hazard of a single event (in absence of competing risks). The practical implications of this for our MI context are that the findings of White and Royston (2009) in the single-event survival setting should in principle extend to the Fine–Gray context. Namely, that the (approximately compatible) directly specified imputation model $f(X_j \mid V, D, X_{-j}, Z; \alpha)$ for a partially observed X_j should contain as predictors at least X_{-j}, Z, the indicator for the competing event of interest $I(D = 1)$, and the cumulative subdistribution baseline hazard for the same event $\Lambda_{01}(V)$, evaluated at the individual subdistribution time V. Instead of the unknown true $\Lambda_{01}(V)$, one could use the estimated marginal cumulative subdistribution hazard $\hat{\Lambda}_1(V)$ instead, obtained using the Nelson–Aalen estimator using V and $I(D = 1)$ are outcome variables. We refer to this approximately compatible MICE approach as FG-Approx.

5.3.3 Accommodating random right-censoring

In addition to (deterministic) administrative censoring, random right-censoring may occur. In the presence of random right-censoring, the contribution of those failing from cause 2 to density Equation 5.6 is no longer evaluable, since we do not know their potential censoring time. Their subdistribution time has effectively been 'coarsened' by their cause 2 failure: we know only that the potential censoring time is later than the cause 2 failure time.

5.3.3.1 Via imputation of potential censoring times

One approach to estimate the parameters of a Fine–Gray model in the presence of random right censoring is to consider the potential censoring times for those failing from cause 2 as missing data, and multiply impute them. To this end, Ruan and Gray (2008) suggested the use of Kaplan–Meier (KM) imputation (Taylor *et al.*, 2002). Specifically, potential censoring times are randomly drawn from the conditional distribution with distribution function $1-P(C > t \mid C > T_i) = 1-\hat{G}(t-)/\hat{G}(T_i-)$, where $\hat{G}(t)$ is a KM estimate of the survival distribution of the censoring times $P(C > t)$ and T_i is the observed event time of an individual failing from a competing event. The imputation of these potential censoring times effectively produces multiple censoring complete datasets, in which a Fine–Gray model can be fit using standard software. Inference is then based on a pooled model, which combines the models fitted in each censoring complete dataset using Rubin's rules (Rubin, 1987).

We can make use of the above ideas in order to multiply impute covariates compatibly with a Fine–Gray model in the presence of random right random censoring. Specifically, we can apply the FG-SMC (or FG-Approx) method in each censoring complete dataset obtained after first imputing the potential censoring times for those failing from cause 2. In order to formalise this procedure, recall that $\boldsymbol{\beta}$ represents the parameters of the substantive model, and that $X = \{X^{\text{obs}}, X^{\text{mis}}\}$. We can similarly partition $V = \{V^{\text{obs}}, V^{\text{mis}}\}$, where V^{mis} is the vector of missing censoring times for those failing from cause 2.

From a Bayesian perspective, the goal is to estimate the conditional density of $\boldsymbol{\beta}$ given the observed data, namely

$$f(\boldsymbol{\beta} \mid X^{\text{obs}}, Z, V^{\text{obs}}, D) = \int_V \int_X f(\boldsymbol{\beta} \mid X^{\text{obs}}, X^{\text{mis}}, Z, V^{\text{obs}}, V^{\text{mis}}, D) \times f(X^{\text{mis}}, V^{\text{mis}} \mid X^{\text{obs}}, Z, V^{\text{obs}}, D) \, \mathrm{d}X^{\text{mis}} \, \mathrm{d}V^{\text{mis}}. \tag{5.7}$$

If we can sample imputed values M times from $f(X^{\text{mis}}, V^{\text{mis}} \mid X^{\text{obs}}, Z, V^{\text{obs}}, D)$, the integral above can be approximated by an average over $f(\boldsymbol{\beta} \mid X^{\text{obs}}, X^{\text{mis}}, Z, V^{\text{obs}}, V^{\text{mis}}, D)$ (the 'complete data' posterior density) evaluated at those M moments (Molenberghs *et al.*, 2014).

One option to sample from $f(X^{\text{mis}}, V^{\text{mis}} \mid X^{\text{obs}}, Z, V^{\text{obs}}, D)$, the joint posterior predictive density, is to use a sequential approach, where we factorise

$$\begin{aligned} f(X^{\text{mis}}, V^{\text{mis}} \mid X^{\text{obs}}, Z, V^{\text{obs}}, D) &= f(X^{\text{mis}} \mid X^{\text{obs}}, Z, V, D) f(V^{\text{mis}} \mid X^{\text{obs}}, Z, V^{\text{obs}}, D), \\ &= f(X^{\text{mis}} \mid X^{\text{obs}}, Z, V, D) f(V^{\text{mis}} \mid Z, V^{\text{obs}}, D). \end{aligned}$$

The above is valid as long as $C \perp X \mid Z$. Practically speaking, this involves imputing the potential censoring times (possibly in strata of Z) in a first step, and then imputing the missing X in a second step. This can be implemented easily using existing software packages in R: {kmi} for the imputation of censoring times (Allignol and Beyersmann, 2010), and {smcfcs} for the imputation of the missing covariates (Bartlett *et al.*, 2022)—see Figure 5.1 for an illustration of the workflow.

If the censoring process additionally depends on the partially observed X, we will need to iteratively sample from $f(V^{\text{mis}} \mid X, Z, V^{\text{obs}}, D)$ and $f(X^{\text{mis}} \mid X^{\text{obs}}, Z, V, D)$ with the following modifications:

1. A model for the censoring process is specified, which must condition X. When this model (which is used to impute the potential censoring times) is fitted during the imputation process, it must be fitted using the most recently imputed values of X. If X is continuous, this could be done using a Cox model for the censoring hazard.

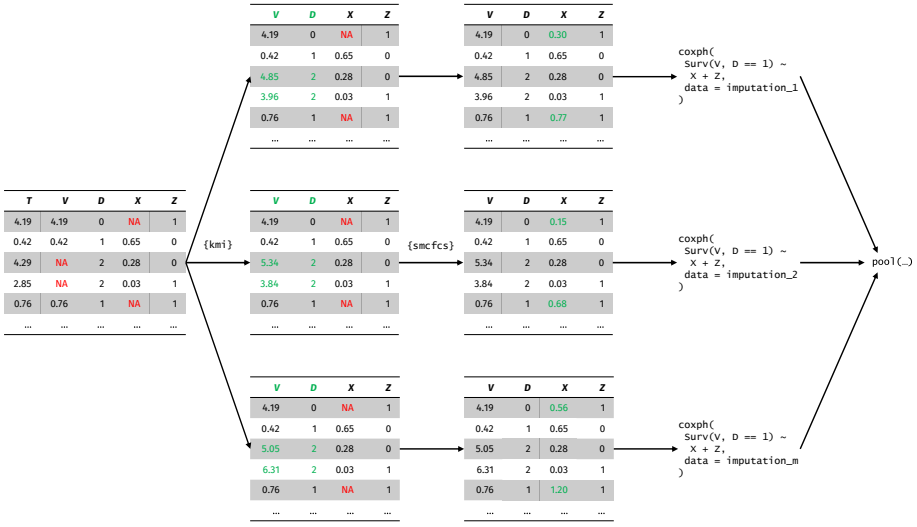

Figure 5.1: Sequential workflow for (compatible) covariate imputation and analysis for a Fine–Gray substantive model with two covariates X and Z, in the presence of random right-censoring. In the first step, the potential censoring times for those failing from cause 2 are multiple imputed using the {kmi} package. In the second step, the missing covariates are imputed using the {smcfcs} (or {mice}) package. This workflow is valid when the probability of being censored is independent of X and any Z related to the censoring process are modelled in {kmi}.

2. Since the censoring distribution partially depends on unobserved values of X, it cannot be ignored in the imputation model. That is, the probability density function of the censoring process no longer factors out of Equation 5.5 (as is the case under random censoring), and is hence non-ignorable or informative (Schluchter and Jackson, 1989). Therefore, the censoring process must be modelled as a cause-specific competing event when imputing the missing X (Borgan and Keogh, 2015). At each iteration, conditional on the most recently imputed potential censoring times, we impute X compatibly with two cause-specific Cox models using CS-SMC: one using V and $I(D = 1)$ as outcome variables (i.e. the Fine–Gray model), and the other using V and $I(D \neq 1)$ (i.e. the model for the censoring process, which must include X).

It is not yet possible to implement the above substantive-model-compatible procedure using existing software in a straightforward way. However, the extension to FG-Approx which accommodates censoring depending on X can be implemented in {mice} using custom imputation methods. The imputation model for X then includes X_{-j}, Z, $I(D = 1)$, $\hat{\Lambda}_1(V)$, and $\hat{H}_C(V)$. Here, $\hat{H}_C(V)$ is the marginal cumulative hazard for the censoring process, estimated based on the most recently imputed V and $I(D \neq 1)$. Based on simulations for other proportional hazards models, we expect that failing to account for non-ignorable censoring at the imputation phase in the present context would have a relatively mild effect on inferences, unless both the proportion of censored observations and the effect of X on the censoring process are large (Bartlett and Taylor, 2016; Borgan and Keogh, 2015; Qi *et al.*, 2010).

Note that the described KM-based procedure for imputing potential censoring times does not take into account any of the uncertainty in estimating $P(C > t)$. That is, the imputed potential censoring times do not involve an initial parameter draw, and are hence not proper Taylor *et al.* (2002). Ruan and Gray (2008) discussed using the non-parametric bootstrap to account for this uncertainty (i.e. each set of imputed censoring times is based on a censoring distribution estimated in a separate bootstrap sample) and improve estimation properties, and found similar results both with and without a bootstrap step. Alternatively, one could choose to specify a (flexible) parametric model for $P(C > t)$, with which we can easily draw from the posterior of all model parameters. In Appendix A, we visualise and give additional details concerning the imputation of potential censoring times.

5.3.3.2 Via censoring weights in the likelihood

Rather than multiply imputing the potential censoring times, an alternative approach is to incorporate inverse probability of censoring weights directly in Equation 5.6. If

we define time-dependent weights

$$w(t) = 1 \qquad \text{if } t \leq T_i,$$
$$w(t) = P(C > t \mid C > T_i) = \frac{G(t-)}{G(T_i-)} \qquad \text{if } t > T_i,$$

then the conditional density of the (subdistribution) outcome given the covariates can be written as

$$f(V, D \mid X, Z) = [\lambda_1(V \mid X, Z) \exp\{-\Lambda_1(V \mid X, Z)\}]^{I(D=1)} \exp\{-\Lambda_1(V \mid X, Z)\}^{I(D=0)} \times \exp\left\{ - \int_0^\infty w(u) \lambda_1(u \mid X, Z) \, \mathrm{d}u \right\}^{I(D=2)}, \tag{5.8}$$

where the term for those failing from the competing event involves integration in practice up to a maximum potential follow-up time t^*. As described by Lambert *et al.* (2017), this integral can be approximated by splitting time into intervals, in which the corresponding $w(t)$ is assumed to be constant.

The integration step needed for those failing from cause 2 in Equation 5.8 means that this approach cannot be implemented in a straightforward way with existing software, unlike the approach described in the previous subsection. The simulation study in this paper therefore focuses on the approach involving multiple imputation of potential censoring times.

5.3.4 Implementation of MI approaches

Methods CS-SMC, CS-Approx, FG-SMC, and FG-Approx can all be implemented using existing software packages in R. In this section, we summarise the steps needed to apply these methods in a given dataset in the presence of random right censoring (possibly in combination with administrative censoring). A minimal R code example can be found in supplementary material S1.

1. Add columns $\hat{H}_1(T)$ and $\hat{H}_2(T)$ to the original data, which are the marginal cause-specific cumulative hazards for each competing risk evaluated at an individual's event or censoring time (obtained using the Nelson–Aalen estimator).
2. Multiply impute the potential censoring for those failing from cause 2 using {kmi}, yielding m censoring complete datasets (i.e. with 'complete' V). The censoring distribution has support at both random and administrative censoring times. Any completely observed covariates that are known to affect the probability of being censored should be included as predictors in the model for the censoring process. {kmi} imputes based on stratified KM when Z are categorical, and based on a Cox model at least one of Z is continuous. If for example an individual's

time of entry into a study determines their maximum follow-up duration, this should be accounted for in the imputation procedure (e.g. by stratifying by year of entry).

3. In each censoring complete dataset, add an additional column $\hat{\Lambda}_1(V)$. This takes the value of the marginal cumulative subdistribution hazard for cause 1 at an individual's observed or imputed subdistribution time, obtained with the Nelson–Aalen estimator based on $I(D = 1)$ and imputed V.

4. In each censoring complete dataset (each with different V and $\hat{\Lambda}_1(V)$, but same $\hat{H}_1(T)$ and $\hat{H}_2(T)$), create a single imputed dataset using the desired covariate imputation method(s):

    - CS-SMC: use {smcfcs} to impute the missing covariate(s) compatibly with cause-specific Cox models. All covariates used in the Fine–Gray substantive model should feature in at least one of the specified cause-specific models.
    - CS-Approx: use {mice} to impute the missing covariate(s), where the imputation model contains as predictors the remaining substantive model covariates, D (as a factor variable), and both $\hat{H}_1(T)$ and $\hat{H}_2(T)$.
    - FG-SMC: use {smcfcs} to impute the missing covariate(s) compatibly with the Fine–Gray substantive model. This is done by using the imputation methods developed for the standard Cox model, but with as outcome variables $I(D = 1)$ and imputed V.
    - FG-Approx: use {mice} to impute the missing covariate(s), where the imputation model contains as predictors the remaining substantive model covariates, $I(D = 1)$, and $\hat{\Lambda}_1(V)$.

5. Fit the Fine–Gray substantive model in each imputed dataset (using standard Cox software with $I(D = 1)$ and imputed V as outcome variables), and pool the estimates using Rubin's rules.

5.4 Simulation study

We aim to evaluate the performance of different MI methods in the presence of missing covariate data when specifying a Fine–Gray model for the subdistribution hazard for one event of interest in the presence of one competing event. Specifically, we assume interest lies in the estimation (for cause 1) of both subdistribution hazard ratios, and the cumulative incidence for a particular individual at some future time horizon. We follow the ADEMP structure for the reporting of the simulation study (Morris *et al.*, 2019).

5.4.1 Data-generating mechanisms

We generate datasets of $n = 2000$ individuals, with two covariates X and Z. We assume $Z \sim \mathcal{N}(0,1)$ and $X \mid Z \sim \text{Bernoulli}\{(1+e^{-Z})^{-1}\}$.

We let $h_k(t \mid X, Z)$, $\lambda_k(t \mid X, Z)$ and $F_k(t \mid X, Z) = P(\tilde{T} \leq t, \tilde{D} = k \mid X, Z)$ respectively denote the cause-specific hazards, subdistribution hazards and cumulative incidence functions for cause k, conditional on X and Z. The competing event times will be generated following two mechanisms: one where the Fine–Gray model for cause 1 is correctly specified, and another where it is misspecified. These are detailed below, together with assumptions concerning both censoring and the missing data mechanisms.

5.4.1.1 Correctly specified Fine–Gray

For this mechanism, we simulate data using the 'indirect' method described in Beyersmann *et al.* (2012), and originally used in the simulations by Fine and Gray (1999). This approach involves first drawing the competing event indicator $\tilde{D}$, and then generating an event time for those with $\tilde{D} = 1$. The final step is to generate times of the competing event for the remaining individuals, who were assigned $\tilde{D} = 2$.

Here, we directly specify the cumulative incidence of cause 1 as

$$F_1(t \mid X, Z) = 1 - \left[1 - p\{1 - \exp(-b_1 t^{a_1})\}\right]^{\exp(\beta_1 X + \beta_2 Z)}.$$

The above expression corresponds to a Fine–Gray model, with as baseline cumulative incidence function a Weibull cumulative distribution function with shape a_1 and rate b_1 (parametrisation used in Klein and Moeschberger, 2003) multiplied by a probability p. Explicitly,

$$F_{01}(t) = p\{1 - \exp(-b_1 t^{a_1})\}.$$

With $\lim_{t\to\infty} F_{01}(t) = p$, we have that $P(\tilde{D} = 1 \mid X, Z) = 1 - (1-p)^{\exp(\beta_1 X + \beta_2 Z)}$, and $P(\tilde{D} = 2 \mid X, Z) = 1 - P(\tilde{D} = 1 \mid X, Z) = (1-p)^{\exp(\beta_1 X + \beta_2 Z)}$. These are the individual-specific cumulative incidences for each event at time infinity. Also note that the baseline subdistribution hazard for this mechanism can be obtained by $\{\mathrm{d}F_{01}(t)/\,\mathrm{d}t\} \times \{1 - F_{01}(t)\}^{-1}$.

The idea then is to generate the event times for cause 1 conditionally on the event indicator and covariates, using

$$\begin{aligned} P(\tilde{T} \leq t \mid \tilde{D} = 1, X, Z) &= \frac{P(\tilde{T} \leq t, \tilde{D} = 1 \mid X, Z)}{P(\tilde{D} = 1 \mid X, Z)} \\ &= \frac{1 - \left[1 - p\{1 - \exp(-b_1 t^{a_1})\}\right]^{\exp(\beta_1 X + \beta_2 Z)}}{1 - (1-p)^{\exp(\beta_1 X + \beta_2 Z)}}. \end{aligned} \tag{5.9}$$

To sample from the above, we first need to draw $\tilde{D} \sim \text{Bernoulli}\{(1-p)^{\exp(\beta_1 X+\beta_2 Z)}\}+1$. We can then use inverse transform sampling to draw failure times within the subset of individuals with $\tilde{D}=1$. Shortening $\exp(\beta_1 X+\beta_2 Z)=\exp(\eta)$, and with $u \sim \mathcal{U}(0,1)$, we can invert Equation 5.9 as

$$t=\left[-\frac{1}{b_1}\log\left[1-\frac{1-\left[1-u\{1-(1-p)^{\exp(\eta)}\}\right]^{1/\exp(\eta)}}{p}\right]\right]^{1/a_1}.$$

For the competing event, we can factorise the cumulative incidence function as

$$P(\tilde{T}\leq t, D=2\,|\,X,Z)=P(\tilde{T}\leq t\,|\,\tilde{D}=2,X,Z)P(\tilde{D}=2\,|\,X,Z).$$

A proportional hazards model can then be specified (for convenience) for

$$P(\tilde{T}\leq t\,|\,\tilde{D}=2,X,Z)=1-\exp\Big\{-H^*_{02}(t)\exp(\beta_1^* X+\beta_2^* Z)\Big\},$$

where $H^*_{02}(t)$ is the cumulative baseline hazard associated to the cumulative incidence function conditional on $\tilde{D}=2$. Since the event indicator is already drawn, the failure times can be drawn again using inverse transform sampling within the subset with $\tilde{D}=2$. Here, we specify a Weibull baseline hazard as $h^*_{02}(t)=a_2 b_2 t^{a_2-1}$.

We fix $\{\beta_1,\beta_2,\beta_1^*,\beta_2^*\}=\{0.75,0.5,0.75,0.5\}$, and the Weibull parameters used for both events as shape $\{a_1,a_2\}=0.75$ and rate $\{b_1,b_2\}=1$. We vary $p=\{0.15,0.65\}$, which is the expected proportion of event 1 failures for individuals with $X=0$ and $Z=0$.

5.4.1.2 Simulation based on cause-specific hazards (misspecified Fine–Gray)

In this data-generating mechanism (DGM), we assume proportionality on the cause-specific hazard scale, and simulate using latent failure times (Beyersmann *et al.*, 2009). We specify baseline Weibull hazards for both cause-specific hazards as

$$h_1(t\,|\,X,Z)=a_1 b_1 t^{a_1-1}\exp(\gamma_{11}X+\gamma_{12}Z),$$
$$h_2(t\,|\,X,Z)=a_2 b_2 t^{a_2-1}\exp(\gamma_{21}X+\gamma_{22}Z),$$

where $\{a_1,a_2\}$ and $\{b_1,b_2\}$ are respectively the shape and rate parameters. Under this DGM, a Fine–Gray model for cause 1 will be misspecified. Nevertheless, the coefficients resulting from the misspecified Fine–Gray model could still be interpreted as time-averaged effects on the (complementary log-log transformed) cumulative incidence function (Grambauer *et al.*, 2010).

We aim to have a scenario close to the one described in Section 5.4.1.1 (in terms of event proportions), where the main difference is that proportionality now holds on the cause-specific hazard scale. To fix the parameters in this DGM, we first simulate a large

dataset of one million individuals following the mechanism described in the previous subsection, where proportional subdistribution hazards hold. Parametric cause-specific proportional hazards models assuming baseline Weibull hazards are then fitted for each failure cause. The point estimates obtained from these models are used as the cause-specific data-generating parameters $\{a_1, b_1, \gamma_{11}, \gamma_{12}\}$ and $\{a_2, b_2, \gamma_{21}, \gamma_{22}\}$. These parameters will of course differ depending on $p = \{0.15, 0.65\}$, and also depending on the censoring distribution. While the cause-specific models fitted on this large dataset will be misspecified (cause-specific baseline hazards are not of Weibull shape, and covariates effects on the cause-specific hazards are non-proportional), the resulting 'least false' parameters are still useful.

Figure 5.2 summarises the DGMs, prior to the addition of any censoring. In the correctly specified Fine–Gray scenarios, the subdistribution log hazard ratio $\lambda_1(t \mid X = 1, Z)/\lambda_1(t \mid X = 0, Z)$ is time constant, while the cause-specific log hazard ratios are time-dependent. The reverse is true for the misspecified Fine–Gray scenarios. Overall, the correctly specified and misspecified Fine–Gray scenarios are very comparable in terms of (true) baseline hazards and cumulative incidences, for both values of p.

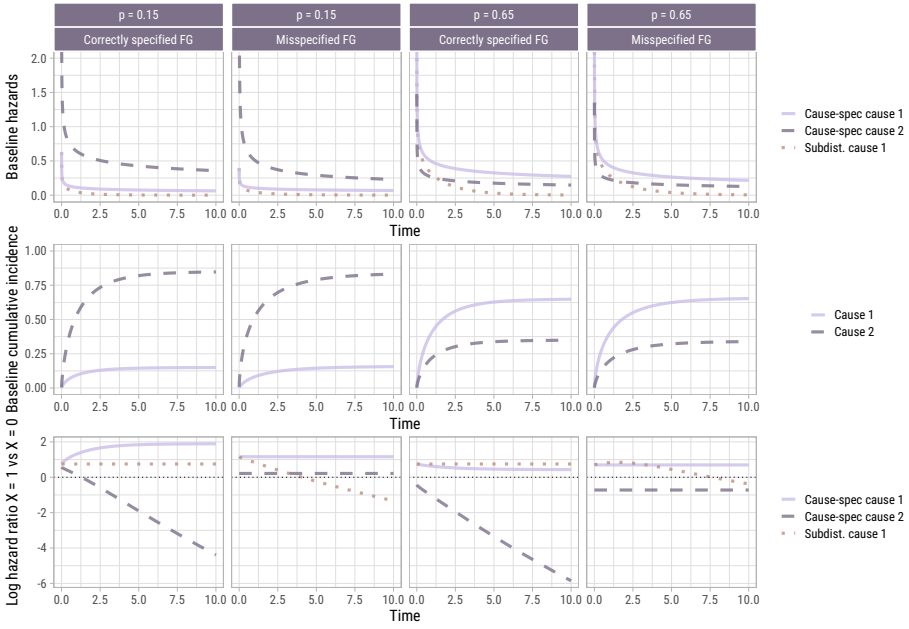

Figure 5.2: Summary of data-generating mechanisms prior to the addition of any censoring. For each value of p, both the correctly specified and misspecified Fine–Gray (FG) scenarios are very comparable in terms of (true) baseline hazards and cumulative incidences.

5.4.1.3 Censoring

The DGMs outlined above assume no loss to follow-up. As additional scenarios, we consider independent (i.e. not conditional on any covariates) right censoring where the censoring times are simulated from an exponential distribution with rate $\lambda_C = 0.49$, resulting in approximately 30% of censored observations. These censoring times will be considered as either: a) known (administrative censoring); or b) unknown (random censoring).

5.4.1.4 Covariate missingness

Missingness is induced in X, while Z remains fully observed. Let R_X be a binary variable indicating whether X is missing ($R_X = 0$) or observed ($R_X = 1$). We use a missing at random (MAR) mechanism conditional on Z, which was defined as $\text{logit}\, P(R_X = 0 \mid Z) = \eta_0 + \eta_1 Z$. We take $\eta_1 = 1.5$, a rather strong mechanism where higher values of Z are associated with more missingness in X. The value of η_0 is found via standard root solving, such that the average probability $P(R_X = 0) = \mathbb{E}\{P(R_X = 0 \mid Z)\}$ of being missing in a given dataset equals 0.4.

5.4.1.5 Summary

In summary, the simulation study varied

- Censoring type: no censoring, administrative and random censoring
- Relative occurrence of event 1, as low or high. This is done by varying the baseline cumulative incidence of event 1 (as $t \to \infty$) as $p = \{0.15, 0.65\}$.
- Failure time simulation methods, with a) directly specified cumulative incidence cause 1 (correctly specified Fine–Gray); b) cause-specific proportional hazards for both causes (misspecified Fine–Gray).

This adds up to 3 (censoring types) × 2 (relative occurence event 1) × 2 (failure time simulation methods) = 12 scenarios.

5.4.2 Estimands

The first estimands of interest are the subdistribution log hazard ratios β_1 and β_2 for X and Z, respectively. In the correctly specified Fine–Gray scenarios, these simply correspond to the data-generating parameters $\{\beta_1, \beta_2\} = \{0.75, 0.5\}$. In the misspecified Fine–Gray scenarios however, the target values (the 'least-false parameters'; time averaged subdistribution log hazard ratios $\{\tilde{\beta}_1, \tilde{\beta}_2\}$) are obtained by fitting a Fine–Gray model on a large simulated dataset of one million individuals, simulated as under the

second data-generating mechanism, after applying any censoring. For computational efficiency, the censoring times are assumed to be known when fitting the Fine–Gray model on this large dataset.

The second estimands of interest are the conditional cumulative incidence of event 1 at a grid of timepoints (between timepoints 0 and 5) for reference individuals $\{X, Z\} = \{0, 0\}$ (baseline) and $\{X, Z\} = \{1, 1\}$. In the correctly specified Fine–Gray scenarios, this corresponds to

$$F_1(t \mid X, Z) = 1 - \left[1 - p\{1 - \exp(-b_1 t^{a_1})\}\right]^{\exp(\beta_1 X + \beta_2 Z)},$$

while for the misspecified Fine–Gray scenarios, this corresponds to

$$\begin{aligned} F_1(t \mid X, Z) &= \int_0^t h_1(u \mid X, Z) \exp\Big\{ -H_1(u \mid X, Z) - H_2(u \mid X, Z) \Big\} \, \mathrm{d}u, \\ &= \int_0^t a_1 b_1 u^{a_1 - 1} \exp(\gamma_{11} X + \gamma_{12} Z) \\ &\qquad \times \exp\Big\{ -b_1 u^{a_1} \exp(\gamma_{11} X + \gamma_{12} Z) - b_2 u^{a_2} \exp(\gamma_{21} X + \gamma_{22} Z) \Big\} \, \mathrm{d}u, \end{aligned}$$

which is obtained via numerical integration.

5.4.3 Methods

The assessed methods are

- Full: analysis run on full data prior to missing values, as a benchmark for the best possible performance.
- CCA: complete-case analysis, as a 'lower' benchmark that the imputation methods need to outperform in order to be worthwhile.
- CS-SMC: MI, imputing compatibly with cause-specific Cox proportional hazards models. This method is described in Section 5.3.1.1. Both X and Z are used as predictors in each cause-specific model assumed by this procedure.
- CS-Approx: MI with both marginal cumulative cause-specific hazards (evaluated at the individual observed event or censoring time) and competing event indicator included as predictors in the imputation model, in addition to Z. This method is described in Section 5.3.1.2.
- FG-SMC: MI, imputing compatibly with a Fine–Gray model for cause 1 that has as covariates X and Z. This is the method described in Section 5.3.2.1.
- FG-Approx: MI with marginal cumulative subdistribution hazard (evaluated at the individual observed or imputed subdistribution time V) and indicator for event 1 included as predictors in the imputation model, in addition to Z. This method is described in Section 5.3.2.2.

The imputation methods are run with 30 imputed datasets. This was fixed following a pilot set of simulations with 50 imputed datasets, which showed that there was little reduction in empirical standard errors for the subdistribution log hazard ratios (and their Monte Carlo standard errors) beyond 30 imputed datasets. Approximately compatible MI methods CS-Approx and FG-Approx only require a single iteration because there is just one variable with missing values, while substantive-model-compatible (SMC) MI methods CS-SMC and FG-SMC are run with 20 iterations. The method used to model $f(X \mid V, D, Z; \alpha)$ for approximately compatible methods is logistic regression, while for SMC methods $f(X \mid Z; \psi)$ is specified as a logistic regression. We note that X was chosen to be binary as SMC methods do not require rejection sampling for variables with discrete sample space, thereby reducing simulation time. If X is chosen to be continuous, the performance of approximately compatible methods is expected to worsen, while no material impact is expected on performance of (correctly specified) SMC methods (Bonneville *et al.*, 2022).

For the scenarios with no or administrative censoring, the subdistribution time V is fully observed. While $V = T$ for those failing from cause 1, for those failing from cause 2, V is first set to either a) a large value greater than the largest observed event 1 time (in absence of censoring); or b) the known potential censoring time C (administrative censoring). The marginal cumulative subdistribution hazard used for the approximate subdistribution MI method is obtained using a marginal model with $I(D = 1)$ and the resulting V as outcome variables. The covariate MI methods are run once these V and $I(D = 1)$ variables have been created. In scenarios with random censoring, the potential censoring times for those failing from cause 2 are multiply imputed using the {kmi} R package with default settings: marginal non-parametric model for the censoring distribution, and no additional bootstrap layer. This yields 30 imputed datasets, each with a different V. In each of these datasets, the marginal cumulative subdistribution hazard is estimated in the same way as described above. Thereafter, the covariate MI methods are run in each of these datasets, yielding one imputed dataset for each imputed V (total of 30 imputed datasets), corresponding to the workflow in Figure 5.1.

For all methods, the Fine–Gray model for cause 1 is estimated using a Cox model with (known or imputed) V and $I(D = 1)$ as outcome variables. When the imputation methods are used (and for all methods when there is random right censoring), the estimated $\hat{\beta}_1$ and $\hat{\beta}_2$ are the results of coefficients pooled using Rubin's rules. Confidence intervals around these estimates are built as described in Section 2.4.2 in van Buuren (2018). For the cumulative incidences, the estimates for the two sets of reference values of X and Z are first made in *each* imputed dataset using Equation 5.1, and thereafter pooled using Rubin's rules after complementary log-log transformation—as described in Morisot *et al.* (2015) and recommended by Marshall *et al.* (2009). This predict-then-pool approach (rather than predicting using a pooled model) has been recommended by multiple authors (Mertens *et al.*, 2020; Wood *et al.*, 2015).

5.4.3.1 Performance measures

The primary measure of interest was bias in the estimated subdistribution log hazard ratios. In order to keep the Monte Carlo standard error (MCSE) of bias under a desired threshold of 0.01, we require $n_{\text{sim}} = 0.2^2/0.01^2 = 400$ replications per scenario, as we expect empirical standard errors to be under 0.2 for all scenarios (based on a pilot run). This number was rounded up to $n_{\text{sim}} = 500$. In addition to bias, we recorded empirical and estimated standard errors, and coverage probabilities. For the cumulative incidence estimates, we focused on both bias and root mean square error (RMSE).

5.4.3.2 Software

Analyses were performed using R version 4.3.1 (R Core Team, 2023). Core packages used were: {survival} version 3.5.7 (Therneau, 2023), {mice} version 3.16.0 (van Buuren and Groothuis-Oudshoorn, 2011), {smcfcs} version 1.7.1 (Bartlett *et al.*, 2022), {kmi} version 0.5.5 (Allignol and Beyersmann, 2010), and {rsimsum} version 0.11.3 (Gasparini, 2018).

5.4.4 Results

We summarise the main findings in this section, with full results available in a markdown file on the Github repository linked at the end of the present manuscript.

5.4.4.1 Subdistribution log hazard ratios

We focus on the results for β_1, together with its time-averaged analogue $\tilde{\beta}_1$ in the scenarios with time-dependent subdistribution hazard ratios. Results concerning bias are summarised in Figure 5.3 for all 12 scenarios, and presented on the relative scale (Monte Carlo standard errors were below the desired 0.01 for both bias and relative bias, across all methods and scenarios).

When the Fine–Gray model for cause 1 was correctly specified, the proposed FG-SMC approach was unbiased regardless of censoring type or (baseline) proportion of cause 1 failures. In contrast, imputing compatibly with the (incorrect) assumption of proportional cause-specific hazards showed strong biases, particularly when $p = 0.15$ in the absence of censoring (25% biased). In the presence of censoring however, this bias dropped to approximately 10%. The CS-Approx method showed consistent downward biases regardless of p and censoring type, while the FG-Approx method was only biased when $p = 0.65$. The latter finding is consistent with previous research in the simple survival setting; namely that the approximately compatible MI approach is expected to work well when cumulative incidence is low (White and Royston, 2009).

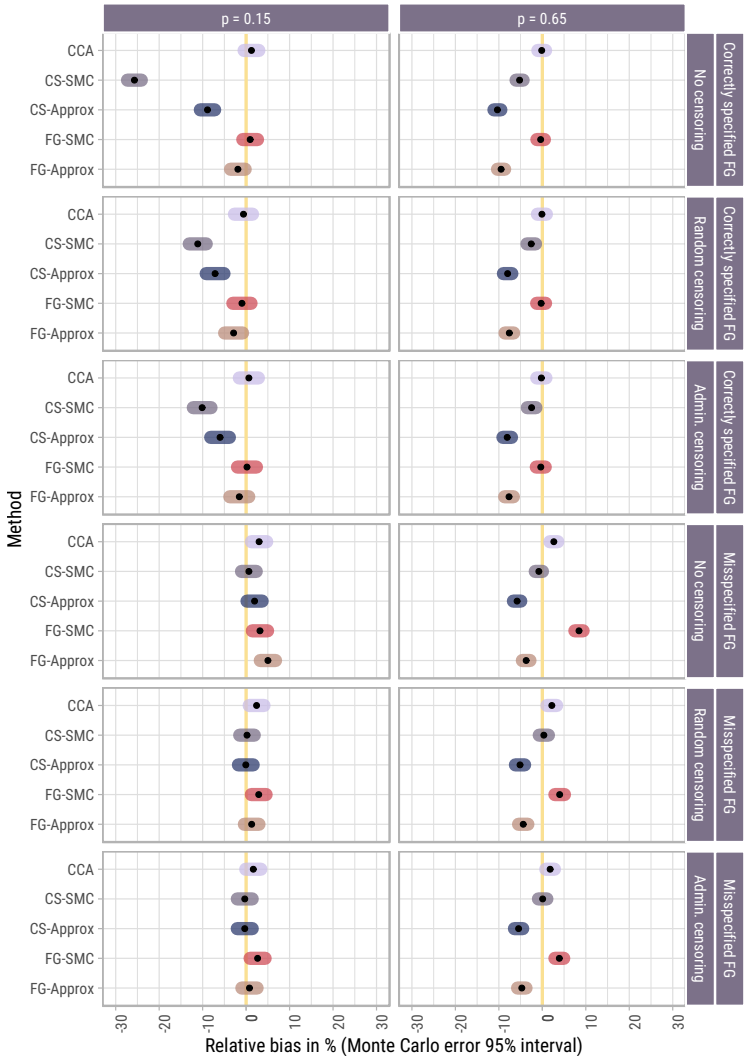

Figure 5.3: Relative bias (%) in estimating β_1, with corresponding 95% Monte Carlo confidence interval (constructed using the standard normal approximation). For each scenario and method, the distribution of $(\hat{\beta}_1 - \beta_1)/\beta_1$ across simulation replications was approximately normal. For the correctly specified Fine–Gray (FG) scenarios, $\beta_1 = 0.75$. In the misspecified FG scenarios, the value of the 'least-false' $\tilde{\beta}_1$ (time-averaged log subdistribution hazard ratio) depended on both p and the presence/absence of censoring. For $p = 0.15$, $\tilde{\beta}_1 \approx 0.76$ without censoring, and $\tilde{\beta}_1 \approx 0.93$ with censoring. For $p = 0.65$, $\tilde{\beta}_1 \approx 0.75$ both with and without censoring.

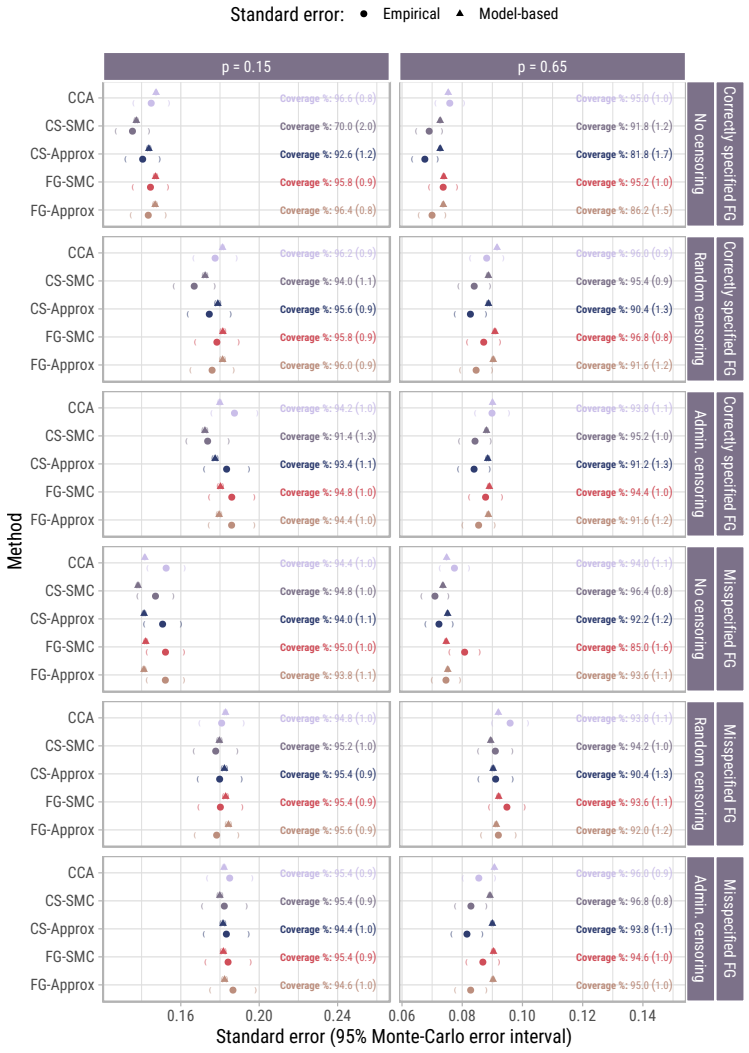

Figure 5.4: Summary of empirical and model-based standard errrors, together with coverage probabilities for β_1 (or $\tilde{\beta}_1$ in scenarios with non-proportional subdistribution hazards). Monte Carlo standard errors (MCSEs) are numerically given in brackets for the coverage probabilities, while the MCSEs for model-based/empirical standard errors are shown by 95% confidence intervals (using standard normal approximation) around a given point. The MCSEs for model-based standard errors are smaller than the graphical width of the point itself, and thus are only visible when zooming-in to the Figure.

When the DGM generated event times under proportional cause-specific hazards, the magnitude of any biases present were in general smaller (e.g. closer to the 5% mark for approximate MI approaches when $p = 0.65$). For the FG-SMC approach, bias was most noticeable when $p = 0.65$, and in the absence of censoring. CS-SMC was unbiased throughout these misspecified Fine–Gray scenarios.

Figure 5.4 summarises empirical and model-based standard errors, together with coverage probabilities for β_1 and $\tilde{\beta}_1$. The model-based standard errors were on average close to their empirical counterparts. CS-SMC appears to have a slight variance advantage over competing approaches, mainly when $p = 0.15$. Interestingly, there was no gain in efficiency when the censoring times were known compared to when they needed to be imputed. This is in line with simulation results in both Fine and Gray (1999) and Ruan and Gray (2008), that compared the censoring complete variance estimator (of subdistribution log hazard ratios) to estimators based on the weighted score function and KM imputation method, respectively. The FG-SMC approach showed good coverage (near the nominal 95% mark) when the Fine–Gray model was correctly specified, although there was slight over-coverage when imputation of censoring times was required. Using the non-parametric bootstrap when estimating $P(C > t)$, which was not investigated in the simulation study, is unlikely to correct for this over-coverage. Under-coverage showed by competing approaches was primarily due to biased estimates.

5.4.4.2 Individual-specific cumulative incidences

Figure 5.5 shows the true and average estimated baseline cumulative incidence function $F_{01}(t)$, the average difference between true and estimated $F_{01}(t)$, and the RMSE of the estimates. Figure 5.6 presents the same information instead for a patient with $\{X, Z\} = \{1, 1\}$. Scenarios where the censoring times are known are omitted from the Figure, as results were indistinguishable from scenarios where the censoring times needed to be imputed.

The cost of imputing compatibly with the wrong model (using CS-SMC when the Fine–Gray model was correctly specified, or FG-SMC when the DGM was based on cause-specific proportional hazards) when estimating $F_{01}(t)$ was only noticeable for the CS-SMC approach in the absence of censoring when $p = 0.15$, in terms of both absolute bias and RMSE. On the whole, the approximately compatible MI approaches performed comparably in terms of RMSE to the SMC approaches. In scenarios where the Fine–Gray model was misspecified, the effect of substantive model misspecification (post-imputation) was clear to see in terms of estimating $F_{01}(t)$ (over- and underestimation at different points in time). In these scenarios, even when CS-SMC is used (which results in the best possible imputations, since it is imputing compatibly with the true data-generating outcome model), assuming proportionality on the incorrect scale at the analysis phase results in biased estimates of the individual-specific cumulative

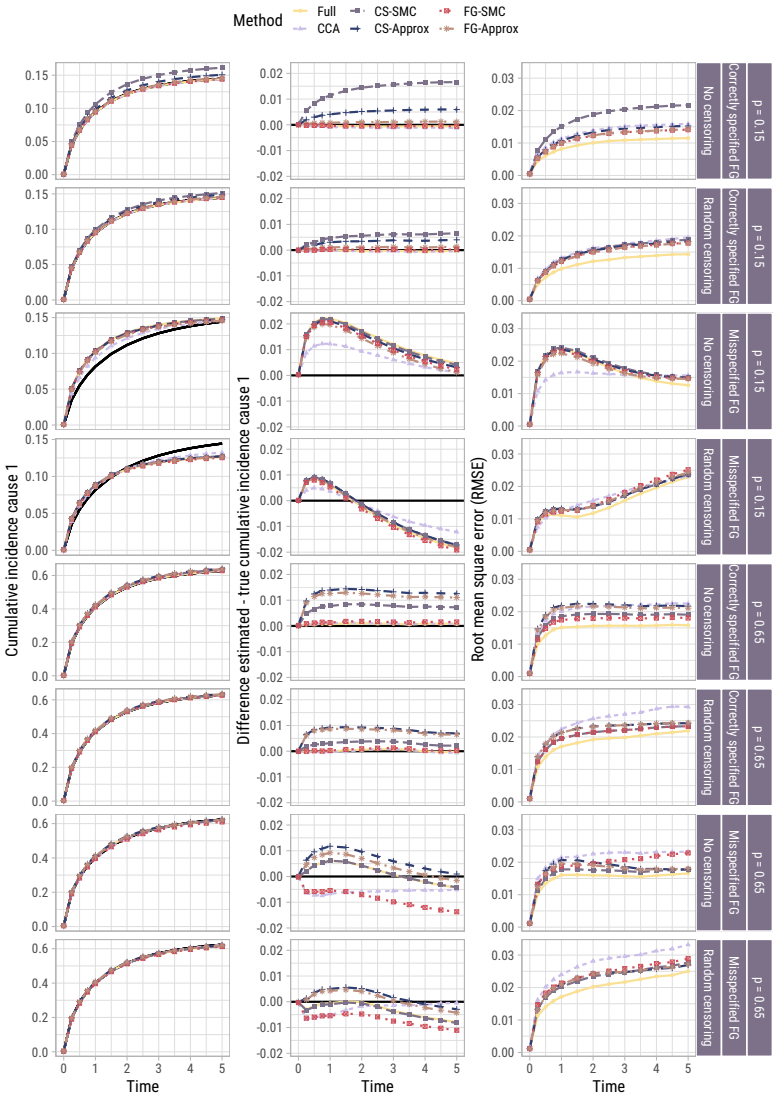

Figure 5.5: Per scenario (row) for a baseline individual $\{X, Z\} = \{0, 0\}$: true (black line) versus estimated cumulative incidence over time, averaged across the 500 replications per scenario (left column); difference between estimated and true (middle column); root mean square error (RMSE) of these estimates (right column). Results for scenarios with administrative censoring are omitted since they were indistinguishable from those with random censoring.

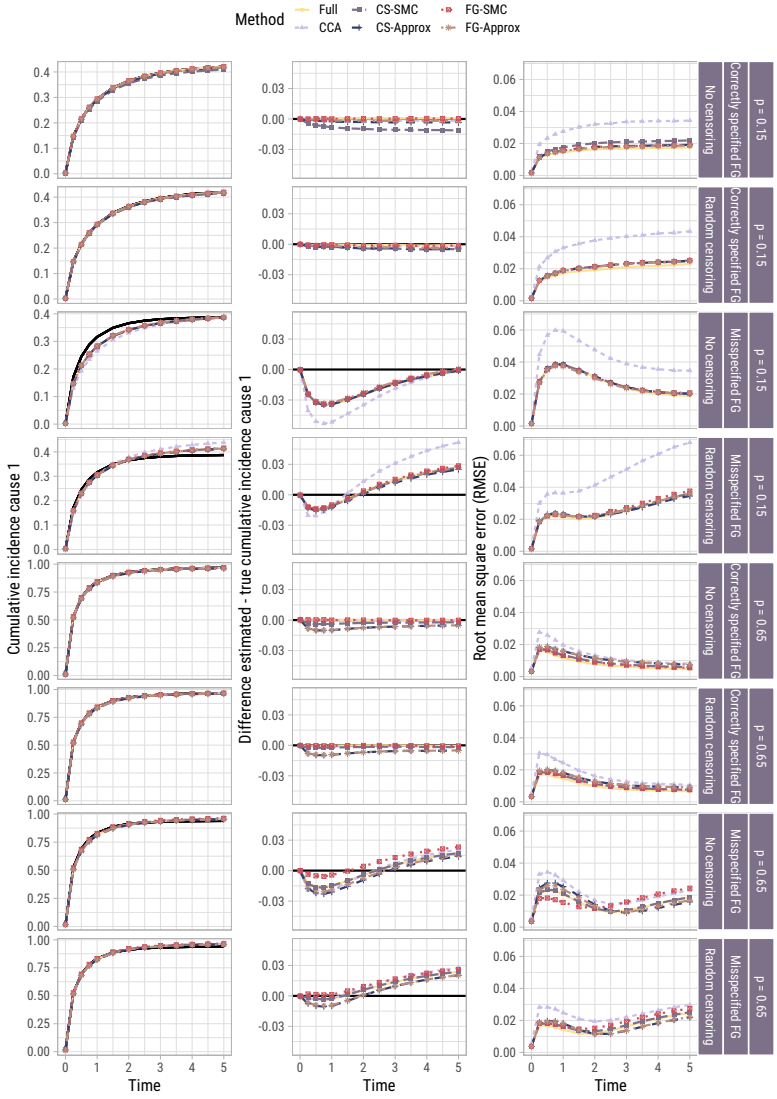

Figure 5.6: Per scenario (row) for individual $\{X, Z\} = \{1, 1\}$: estimated versus true (black line) cumulative incidence over time, averaged across the 500 replications per scenario (left column); difference between estimated and true (middle column); root mean square error (RMSE) of these estimates (right column). Results for scenarios with administrative censoring are omitted since they were indistinguishable from those with random censoring.

incidence function. When $\{X, Z\} = \{1, 1\}$, all imputation approaches outperformed CCA in terms of RMSE when estimating $F_1(t \mid X = 1, Z = 1)$, though to a lesser extent when $p = 0.65$. This can presumably be attributed to the efficiency gain in estimating β_2.

5.4.5 Additional simulations

Two additional sets of simulations were conducted, which build upon the correctly specified Fine–Gray data-generating mechanism with random censoring, with both $p = 0.15$ and $p = 0.65$. The objectives of these additional simulations were to assess the performance of the different imputation methods in settings where a) missingness depends on the observed competing risks outcomes; b) censoring depends on complete covariates, and the model used to impute the potential censoring times could potentially be misspecified.

Covariate imputation approaches were used as previously described in Section 5.4.3, and similarly these additional scenarios are each comprised of 500 simulation replications. The results of these simulations are presented in Appendix B, with a focus on the relative bias in estimating both β_1 and β_2, and further described below.

5.4.5.1 Outcome-dependent missingness

In the first set of simulations, the missingness in X was made to depend on the observed event time T as $\operatorname{logit} P(R_X = 0 \mid T) = \eta_0 + \eta_1 \log(T + 1)$, with $\eta_1 = -1.5$ and η_0 chosen such that 40% of observations in X are missing. This reflects a setting where baseline variables such as genetic information are retrospectively ascertained, and more likely to be available the longer an individual is in follow-up. Since this missingness mechanism depends partially on the failure times for those failing from cause 2, these simulations allow to assess the violation of the MAR assumption made by both FG-SMC and FG-Approx—see Section 5.3.2.1.

To briefly summarise, in Figure 5.9 (Appendix B), we see that the violation of the MAR assumption led to appreciable biases in the estimation of subdistribution log hazard ratios when the proportion of competing events was large (i.e. under $p = 0.15$). In this same scenario, CS-SMC outperformed other methods since it conditions also on the failure time from cause 2, but was still biased as it is imputing compatibly with cause-specific Cox models, which are the incorrect underlying outcome model. CCA was expectedly biased in these scenarios as missingness depended on the outcome.

5.4.5.2 Covariate-dependent censoring

In the second set of simulations, exponential censoring was made covariate-dependent with rate $\lambda_C = 0.49e^Z$, which yields an average censoring proportion which is comparable to the previously reported scenarios with random censoring (approximately 30% censored). In these scenarios, all covariate imputation approaches were applied after multiply imputing the potential censoring times using either a) a marginal KM estimate of the censoring distribution (misspecified censoring distribution); b) a Cox model for the censoring distribution, conditional on Z (correctly specified censoring distribution). The missingness in X also depended on Z, as outlined in Section 5.4.1.4.

In Figure 5.10 (Appendix B), we see that incorrectly specifying the model for the censoring distribution under covariate-dependent censoring led to large biases in the estimation of the subdistribution log hazard ratio for the variable related to the censoring mechanism (β_2 in these simulations). These biases were far less severe under $p = 0.65$, since there are fewer censoring times to impute. Interestingly, in these scenarios CS-SMC does not appear to pay a price for imputing compatibly with the incorrect underlying outcome model. FG-SMC was unbiased throughout when the model for the censoring was correctly specified.

5.5 Applied data example

We illustrate the methods assessed in the simulations study on a dataset of 3982 adult patients with primary and secondary myelofibrosis undergoing a hematopoietic stem cell transplantation (alloHCT) between 2009 and 2019, and registered with the European Society for Blood and Marrow Transplantation (EBMT) (Polverelli *et al.*, 2024). Myelofibrosis is a rare and chronic myeloproliferative neoplasm characterised by bone marrow fibrosis and extramedullary hematopoiesis, for which an alloHCT is the only treatment that can offer long term remission (Kröger *et al.*, 2024). In the original study, the primary objective was to evaluate the association between comorbidities at time of alloHCT and (cause-specific) death without prior relapse of the underlying disease, the so-called non-relapse mortality. In the present illustration, we instead assume that interest lies in developing a prognostic model for time to disease relapse in the first 60 months following an alloHCT. To this end, we developed a Fine–Gray model for relapse, with death prior to relapse as sole competing risk.

A set of 18 baseline predictors were chosen on the basis of substantive clinical knowledge, many of which had a considerable proportion of missing data (see Table 5.1 in Appendix C). These predictors included the 13 variables used in the multivariable models from the original study, and 5 additional variables that were either known to be predictive of disease relapse (use of T-cell depletion; presence of cytogenetic abnormalities), or provided relevant auxiliary information regarding the missing values

(year of transplantation; time between diagnosis and transplantation; and whether diagnosis was primary or secondary myelofibrosis). Note that since this is a model for (complementary log-log transformed) cumulative incidence of relapse, we want to make sure to include predictors known to be associated with the cause-specific hazards of *both* relapse and non-relapse mortality.

Since around 45% of patients were either event-free or censored within the first 60 months (see supplementary material S2.2, non-parametric curves), potential censoring times for those experiencing non-relapse mortality were first multiply imputed using the {kmi} package in strata defined by (completely observed) year of transplantation, yielding 100 datasets with 'complete' subdistribution time V but with partially observed covariate information. In each of these datasets, covariates were imputed once using each of the four imputation methods used in the simulation study, after 20 cycles across the covariates. The choice of 100 imputed datasets was motivated using von Hippel's quadratic rule (i.e. number of imputed datasets needed should increase approximately quadratically with increasing fraction of missing information), based on an initial set of 30 imputed datasets (von Hippel, 2020). Essentially, we sought to control the MCSEs of the standard errors of the estimated subdistribution log hazard ratios. Default imputation methods were used depending on the type of covariate: binary covariates using logistic regression, ordered categorical using proportional odds regression and nominal categorical using multinomial logistic regression. For continuous covariates, the default in {mice} is predictive mean matching, while linear regression is used for $f(X_j \mid X_{-j}, Z; \psi)$ in {smcfcs}. The imputation model for a given partially observed variable therefore contained as predictors all remaining fully and partially observed variables from the substantive model, together with the outcome. Each imputation approach differs mainly in how they incorporate the outcome in the imputation model: either by sampling directly from an assumed substantive model compatible distribution (FG-SMC and CS-SMC), or by including event indicator(s) and marginal cause-specific or subdistribution cumulative hazard(s) explicitly as additional predictors (FG-Approx and CS-Approx).

Figure 5.7 shows for all methods the estimated baseline cumulative incidence function, and the width of the corresponding confidence interval at each timepoint. As was the case in the simulation study, cumulative incidences are estimated in each imputed dataset, and pooled after complementary log-log transformation. The estimation procedure used for the standard errors of the cumulative incidences is described by Ozenne *et al.* (2017). The estimates using both FG-SMC and FG-Approx are virtually overlapping, which is consistent with the simulation study results when $p = 0.15$. Both CS-SMC and CS-Approx also yielded cumulative incidences that were close to those obtained by the subdistribution hazard based imputation approaches, which is in line with the results of the simulation study under random right censoring. The most stark differences were between CCA (which only uses 20% of patients) and the imputation approaches: the cumulative incidence of relapse at 60 months was almost 5% lower than the nearest MI-based curve, with confidence intervals that

were over twice as wide. For completeness, in supplementary material S2.3 we report the pooled subdistribution log hazard ratios, in addition to the pooled coefficients of cause-specific Cox models for relapse and non-relapse mortality (each containing the same predictors as the Fine–Gray model for relapse). The pooled coefficients of the Fine–Gray models were extremely similar between imputation approaches, and all differed considerably from the (much more variable) CCA. There were some noticeable differences between subdistribution hazard based and cause-specific hazard based imputation approaches when estimating the cause-specific Cox model for non-relapse mortality (see e.g. pooled coefficients for weight loss prior to transplantation, hemoglobin or high risk comorbidity score). Furthermore, the pooled subdistribution log hazard ratios were generally small in magnitude (none exceeding 0.5), a setting in which both SMC and approximately compatible approaches are expected to perform similarly.

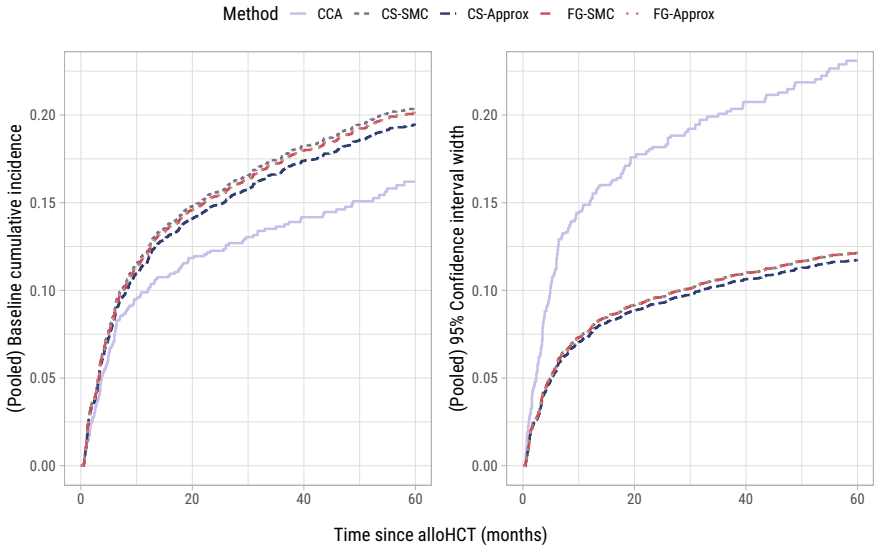

Figure 5.7: Pooled baseline cumulative incidence functions for relapse in the applied data example (left panel), and width of corresponding confidence intervals (right panel). These are the estimates for a patient aged 60, transplanted in 2019 immediately after diagnosis, with 10g/dL hemoglobin, 15 x10^9/L white blood cells, no peripheral blood blasts, and with reference levels for all categorical predictors (see Table 5.1).

The differences observed between point estimates obtained using the imputation based approaches and CCA are in large part explainable by the gulf in efficiency between the two approaches. Nevertheless, there are indications that the estimates obtained

using imputation methods could be less biased than their CCA counterparts in this example. An exploratory logistic model showed that the observed time to competing event and competing event indicator were both predictive of the probability of being a complete-case, after adjusting for other known important predictors of missingness such as year of transplantation (many variables recorded more often later on in time as their clinical relevance became clearer). Upon closer inspection, it appears that the probability of being a complete-case is significantly lower only for those censored earlier on in time. This seemingly unlikely association between future outcome and baseline complete-case indicator (outcome-dependent MAR, under which CCA is biased unless the missingness is related solely to the censoring process—see Rathouz, 2007) is likely confounded by transplant centre. That is, shorter follow-up times and missing values in covariates may both be symptomatic of a given centre's overall quality of data collection. Although ignored in the present analysis for simplicity, there is indeed heterogeneity in data completeness between EBMT affiliated transplant centres across and within different countries. The MI of potential censoring times would allow to model centre effects using standard software, for example by means of stratification or use of a frailty term.

5.6 Discussion

In this paper, we extended the SMC-FCS approach in order to impute missing covariates compatibly with a Fine–Gray substantive model. For a given competing event, the theory relies on using the subdistribution time V and the corresponding event-specific indicator as outcome variables. In the presence of random right-censoring, V is only partially observed, as the potential censoring times for those failing from competing events are unknown. These can be multiply imputed in a first step, after which covariates can be imputed by conditioning on the 'complete' outcome variables. The approach is straightforward to implement in R by making use of existing software packages {kmi} and {smcfcs}. While the imputation of potential censoring times appears underused in the subdistribution hazard modelling literature (relative to weighted approaches), it has inspired other methodological extensions e.g. enabling the use of deep learning in discrete time after single imputation of potential censoring times (Gorgi Zadeh *et al.*, 2022).

The simulation study compared the performance of the proposed method to competing MI approaches, including imputing compatibly with cause-specific proportional hazards models. The FG-SMC approach performed optimally (in terms of estimating both subdistribution log hazard ratios, and cumulative incidences) when the assumption of proportional subdistribution hazards held, and performed satisfactorily when this assumption did not hold. For cumulative incidence estimation, the choice of substantive model (i.e. cause-specific Cox vs. Fine–Gray) at the analysis phase appears

to be more important than the procedure used to impute the missing covariates. In terms of RMSE of these predictions, most imputation approaches outperform CCA. The applied data example also demonstrated the possible gain in efficiency when using MI instead of CCA.

One counterintuitive finding was that the presence of censoring seems to *improve* the performance of the misspecified SMC-FCS procedure (e.g. use of CS-SMC when underlying DGM assumes proportional subdistribution hazards). An explanation for this phenomenon is that the time-dependent factor relating the cause-specific and subdistribution hazards for cause 1 (the 'reduction factor,' Putter *et al.*, 2020) is closer to 1 earlier in time. Therefore (in the example with DGM assuming proportional subdistribution hazards), the violation of proportionality on the cause-specific hazard scale will appear to be less severe in earlier time-periods, thereby improving the performance of the misspecified SMC-FCS approach. This is also in line with earlier findings showing how similar the results of subdistribution and cause-specific hazards models can be in presence of heavy censoring (Grambauer *et al.*, 2010; van der Pas *et al.*, 2018). Notwithstanding, the additional simulations in Section 5.4.5 emphasise the importance of appropriately accounting for covariates related to the censoring process when modelling the subdistribution hazard (where in practice, completely random censoring is the default assumption—see Beyersmann *et al.*, 2012), as also discussed in previous work (Donoghoe and Gebski, 2017).

An advantage to the proposed approach is that it can be extended in various ways. For example, the approach can account for time-dependent effects, by making direct use of existing approaches developed in the context of standard Cox models (Keogh and Morris, 2018). Additionally, the proposed approach can be extended to accommodate interval censored outcomes, using the methodology described by Delord and Génin (2016), which relies on analogous principles: multiply impute interval censored V in order to work with simpler censoring complete data.

There are multiple limitations to the present work. The first is that the proposed SMC-FCS approach does not accommodate delayed entry (left truncation). Our current recommendation to impute approximately compatibly with a Fine–Gray model subject to delayed entry and right-censoring is to include $I(D = 1)$ and $\hat{\Lambda}_1(T)$ as predictors in the imputation model, in addition to other substantive model covariates. Here, $\hat{\Lambda}_1(t)$ is the estimated cumulative subdistribution hazard based on a marginal model that uses time-dependent weights in order to accommodate both left-truncation and right-censoring (Geskus, 2011). Note the proposed imputation model uses $\hat{\Lambda}_1(T)$ and not $\hat{\Lambda}_1(V)$, and therefore some downward bias is to be expected, as explained in Appendix A. Second, while FG-SMC does not require an explicit model for the competing risks, it does require the censoring distribution to be specified explicitly (e.g. non-parametrically using KM, or using a Cox model). Third, the proposed approach is geared towards imputing missing covariates when only one competing event is of interest. More generally, the strategy of estimating a Fine–Gray for each cause in turn

is not an approach the current authors endorse, based on both theoretical (Austin *et al.*, 2021; Beyersmann *et al.*, 2012) and simulation-based arguments (Bonneville *et al.*, 2024). When multiple competing events are of interest, we would instead recommend modelling the cause-specific hazards, or using the semiparametric approach suggested by Mao and Lin (2017) for joint inference on the cumulative incidence functions.

In conclusion, the proposed approach is most appropriate for imputing missing covariates in the context of prognostic modelling of only one event of interest. Based on the simulation study, imputing compatibly with cause-specific proportional hazards seems to be a good all-round strategy for a 'complete' competing risks analysis (investigating both the cause-specific hazards and cumulative incidence functions, see Latouche *et al.*, 2013), and can at the same time be used for prognostic modelling based on the cause-specific Cox models.

Supplementary materials

Supplementary materials for this work are available at https://arxiv.org/abs/2405.16602.

All R code (needed to reproduce simulation study, applied data example, and manuscript figures) is available at https://github.com/survival-lumc/FineGrayCovarMI. In addition to the minimal R code provided in the supplementary materials, a wrapper function for the proposed SMC-FCS Fine–Gray method is available inside the {smcfcs} R package.

Appendix A: Imputed censoring times, and resulting cumulative subdistribution hazards

As described in Section Section 5.3.3.1, the subdistribution time V is only partially observed in the presence of random right-censoring. Thus, the potential censoring times for those failing from cause 2 should first be multiply imputed, before imputing any missing covariates. This imputation of partially observed V is visualised more closely in Figure 5.8, using a simulated dataset of 2000 individuals following the parametrisation used in the simulation study scenario with correctly specified Fine–Gray, $p = 0.65$, and random exponential censoring. In this example, the potential censoring times for those failing from cause 2 were imputed $m = 10$ times.

The upper panel shows the imputed potential censoring times for a random selection of 20 individuals failing from cause 2, in addition to their cause 2 failure time and their true eventual censoring time. The lower panel shows the estimated marginal

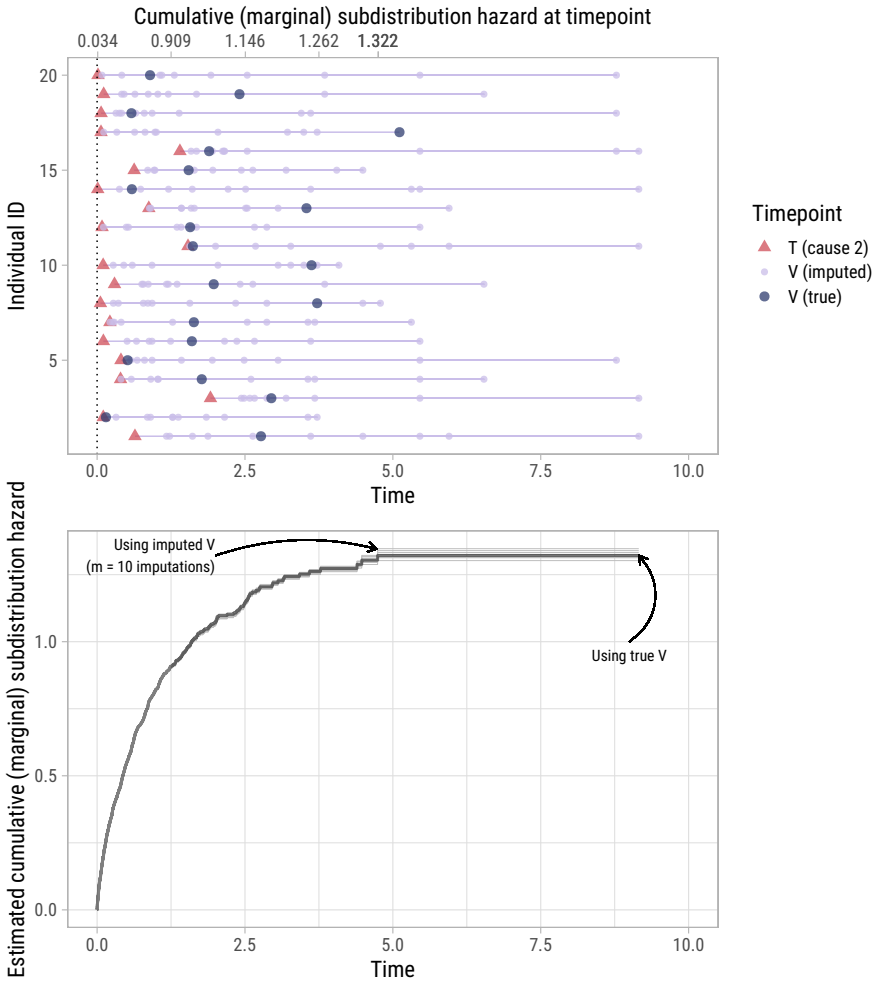

Figure 5.8: Based on a simulated dataset of $n = 2000$ (correctly specified Fine–Gray, $p = 0.65$, random censoring), we show the imputed ($m = 10$ imputations) potential censoring times for a random selection of 20 individuals failing from cause 2 (upper panel); and the estimated marginal cumulative subdistribution hazard function for cause 1 based on true V, and based on imputed V (lower panel).

cumulative subdistribution hazard function for $\hat{\Lambda}_1(t)$ resulting from using $I(D = 1)$ together with either the imputed or true V as outcomes in a marginal model. We used $\hat{\Lambda}_1(t)$ estimated using the true V to create the secondary x-axis in the upper panel, which shows the value of this function at a given timepoint. For example, the marginal cumulative subdistribution hazard was 1.146 at timepoint 2.5, and stayed constant at 1.322 after the last cause 1 event in this sample.

The upper panel in particular gives additional insights regarding the FG-Approx method, where $I(D = 1)$ and $\hat{\Lambda}_1(V)$ are included as predictors in the imputation model. Namely, the secondary x-axis shows the value of $\hat{\Lambda}_1(V)$ used in the imputation model for a missing X_j, for given imputed V. A first key point is that one should always use $\hat{\Lambda}_1(V)$ in the imputation model, and not $\hat{\Lambda}_1(T)$. Since the observed cause 2 failure time occurs before the eventual censoring time, $\hat{\Lambda}_1(T)$ will always be smaller than the marginal cumulative subdistribution hazard at the eventual censoring time. Since $\hat{\Lambda}_1(T)$ and $\hat{\Lambda}_1(V)$ are not proportional to each other, the imputation model will incur some bias. A second point is that in settings with fewer event 1 failures (e.g. $p = 0.15$ scenario in the simulation study), the corresponding secondary x-axis will have a smaller range, since the subdistribution hazard will be lower overall. Using $\hat{\Lambda}_1(T)$ instead of $\hat{\Lambda}_1(V)$ may therefore have a more limited impact. However, as evidenced by the simulations in Section 5.4.5, misspecification the censoring distribution impacts inferences more when $p = 0.15$, since there are more censoring times to impute.

The lower panel shows that the estimated $\hat{\Lambda}_1(t)$ varies very little between imputed datasets, with differences only being noticeable later on in follow-up as risk sets become smaller and associated cumulative hazard jumps more pronounced. Note also that while $\hat{\Lambda}_1(t)$ based on the true V appears in this dataset to be a kind of 'average' of the functions based on imputed V, this will not be the case in general, especially with smaller sample sizes. The $\hat{\Lambda}_1(t)$ based on the weighted estimator (Geskus, 2011) will however coincide with the 'average' of the functions based on imputed V, as will using the negative log of one minus the Aalen–Johansen estimate of the marginal cumulative incidence function.

Appendix B: Additional simulations

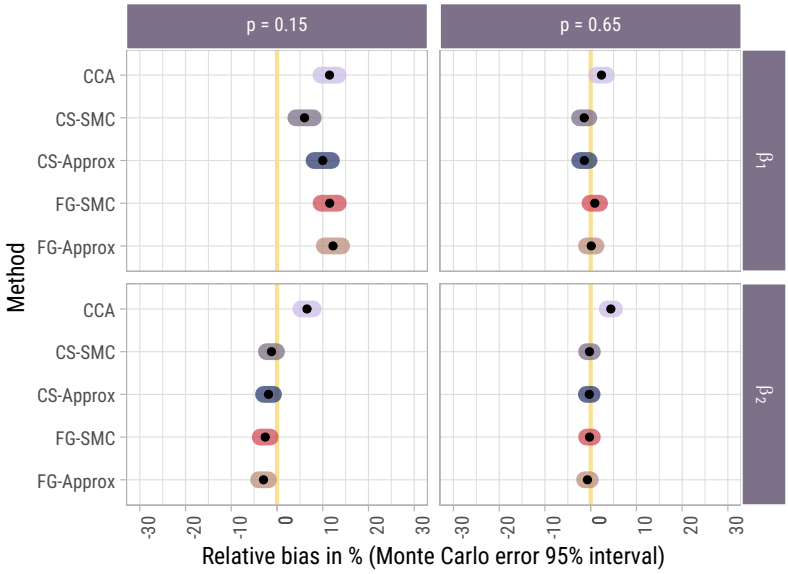

Figure 5.9: Relative bias (%) in estimating $\{\beta_1, \beta_2\} = \{0.75, 0.5\}$, with corresponding 95% Monte Carlo confidence interval (constructed using the standard normal approximation). These are additional simulations under the correctly specified Fine–Gray data-generating mechanism with random censoring, with both $p = 0.15$ and $p = 0.65$. The missingness in X was made to depend on the observed event time T as $\text{logit}\, P(R_X = 0 \mid T) = \eta_0 + \eta_1 \log(T + 1)$, with $\eta_1 = -1.5$ and η_0 chosen such that 40% of observations in X are missing.

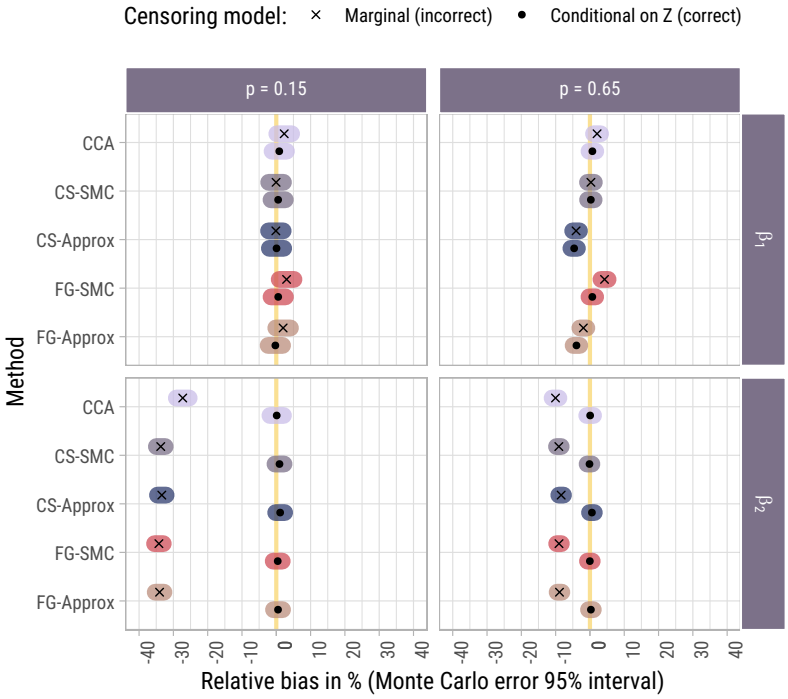

Figure 5.10: Relative bias (%) in estimating $\{\beta_1, \beta_2\} = \{0.75, 0.5\}$, with corresponding 95% Monte Carlo confidence interval (constructed using the standard normal approximation). These are additional simulations under the correctly specified Fine–Gray data-generating mechanism with random censoring, with both $p = 0.15$ and $p = 0.65$. The censoring was made covariate-dependent with rate $\lambda_C = 0.49e^Z$, and all covariate imputation approaches were applied after multiply imputing the potential censoring times using either a) a marginal (incorrect) Kaplan–Meier estimate of the censoring distribution; b) a Cox model for the censoring distribution, conditional on Z (correct). The missingness in X here also depended on Z.

Appendix C: Data dictionary

Table 5.1: Data dictionary. The median and interquartilerange are reported for continuous variables, and for categorical variables the counts and proportion per level are reported. The difference between this cohort (n = 3982) and the one reported in Chapter 4 (n = 4086), is that the present cohort is selected on available outcome data (i.e. no missing relapse times). CMV: cytomegalovirus; HLA: human leukocyte antigen; HCT-CI: Hematopoietic stem cell transplantation-comorbidity index; MF: myelofibrosis.

Variable or level	N = 3982
Patient age (years)	58 (52, 64)
Patient/donor CMV match	
Patient negative/Donor negative	1,142 (30%)
Other	2,715 (70%)
(Missing)	125
Donor type	
HLA identical sibling	1,183 (30%)
Other	2,795 (70%)
(Missing)	4
Hemoglobin (g/dL)	9.10 (8.10, 10.40)
(Missing)	1,873
HCT-CI risk category	
Low risk (0)	1,674 (54%)
Intermediate risk (1 – 2)	743 (24%)
High risk (≥ 3)	674 (22%)
(Missing)	891
Interval diagnosis-transplantation (years)	3 (1, 9)
Karnosfky performance score	
≥ 90	2,475 (66%)
80	986 (26%)
≤ 70	267 (7.2%)
(Missing)	254
Patient sex	
Female	1,484 (37%)
Male	2,498 (63%)
Peripheral blood (PB) blasts (%)	1.0 (0.0, 3.0)
(Missing)	2,323
Conditioning	
Standard	1,373 (35%)
Reduced	2,553 (65%)
(Missing)	56
Ruxolitinib given	
No	1,832 (66%)
Yes	931 (34%)
(Missing)	1,219
Disease subclassification	
Primary MF	2,912 (73%)
Secondary MF	1,070 (27%)
Night sweats	
No	1,256 (70%)
Yes	529 (30%)

(Missing)	2,197
T-cell depletion (in- or ex-vivo)	
No	1,012 (26%)
Yes	2,905 (74%)
(Missing)	65
Cytogenetics	
Normal	1,318 (59%)
Abnormal	910 (41%)
(Missing)	1,754
White blood cell count (WBC, $x10^9$/L)	7 (4, 14)
(Missing)	1,884
>10% Weight loss prior to transplantation	
No	1,329 (73%)
Yes	492 (27%)
(Missing)	2,161
Year of transplantation	2,015.0 (2,012.0, 2,018.0)