



# Imputation of Missing Data for Time-to-Event Endpoints Using Retrieved Dropouts

Shuai Wang<sup>1</sup> · Robert Frederick<sup>2</sup> · James P. Mancuso<sup>3</sup>

Received: 16 April 2023 / Accepted: 21 August 2023 / Published online: 7 October 2023  
© The Author(s) 2023

## Abstract

We have explored several statistical approaches to impute missing time-to-event data that arise from outcome trials with relatively long follow-up periods. Aligning with the primary estimand, such analyses evaluate the robustness of results by imposing an assumption different from censoring at random (CAR). Although there have been debates over which assumption and which method is more appropriate to be applied to the imputation, we propose to use the collection of retrieved dropouts as the basis of missing data imputation. As retrieved dropouts share a similar disposition, such as treatment discontinuation, with subjects who have missing data, they can reasonably be assumed to characterize the distribution of time-to-event among subjects with missing data. In terms of computational intensity and robustness to violation of underlying distributional assumption, we have compared parametric approaches via MCMC or MLE multivariate sampling procedures to a non-parametric bootstrap approach with respect to baseline hazard function. Each of these approaches follows a process of multiple imputation (“proper imputations”), analysis of complete datasets, and final combination. The type-I error, and power rates are examined under a wide range of scenarios to inform the performance characteristics. A subset of a real unblinded phase III CVOT is used to demonstrate the application of the proposed approaches, compared to the Cox proportional hazards model and jump-to-reference multiple imputation.

**Keywords** Time to event · Missing data · Multiple imputation · Cardiovascular outcome trial · Estimand · Informative censoring · Proper imputation

## Introduction

In the 2008 published regulatory guidance [1] for type 2 diabetes medications, cardiovascular outcome trials (CVOTs) were required to demonstrate that the investigational drug would not increase the risk of major adverse cardiovascular events by 30% or more, compared to placebo. All subsequent CVOTs in this field to date have met this requirement and the regulatory agency is in the process of rolling out a new guidance [2] to replace the 2008 version, with a new focus on size of the safety trial and patient population characteristics. These safety trials feature very large sample size, long duration and consequently come at very high costs, aiming to

rule out cardiovascular risk increase and potentially to show cardiovascular or renal benefits. Beyond diabetes, CVOTs are also commonly conducted in lipid-lowering therapies [3, pp. 1713–1722, 4, pp. 2387–2397, 5, pp. 1527–1539]. Powered by even larger sample size and higher number of events to accrue, CVOTs for lipid-lowering drugs usually aim to demonstrate cardiovascular benefits.

In most CVOTs, time to event endpoints are primarily analyzed using a semi-parametric Cox proportional hazards (CPH) model in which discontinued subjects with no prior events are censored at their study discontinuation date, assuming non-informative censoring [6, pp. 139–156], usually justifiable by censoring at random in each treatment group [7, pp. 1–6, 8]. Although every study aims to keep all participants in the study, some subjects will prematurely discontinue study for various reasons including withdrawal of consent, lost to follow-up, physician/sponsor decision, subject move, site closure, etc.

Missing data can be generally defined as subjects who prematurely discontinue the study with no prior events that

✉ Shuai Wang  
shuai1107@hotmail.com

<sup>1</sup> Pfizer Inc., 1 Portland St, Cambridge, MA 02139, USA

<sup>2</sup> Pfizer Inc., 66 Hudson Blvd, New York, NY 10001, USA

<sup>3</sup> Pfizer Inc., 280 Shennecossett Rd, Groton, CT 06340, USA

can be counted towards the endpoint of interest. If the endpoint of interest has already been reached, followed by the subject's discontinuation from the study, then that would not be considered missing data according to this definition. As most CVOTs aim to follow subjects from randomization until death or end of study, whichever occurs first, deaths occurring during the study will generally not be considered as missing data. For example, consider a scenario in which a subject dies during the study and the endpoint is time to first occurrence of 3-point MACE (defined as a composite of non-fatal MI, non-fatal stroke and CV death). If this subject has died due to CV cause, he/she will not be counted as missing data because his/her CV death is counted as a MACE event. In the same example, if he/she dies due to non-CV cause and he/she has no MACE events prior to death, he/she will be counted as complete data with no event, and a time-to-event will not be imputed. Overall, it's not trivial to handle death data because death events often come with limited data on which to adjudicate the cause. Because missing data censored at random in CPH model can potentially result in a biased estimate [9, p. 4, 10, p. 40], we take a missing not at random (MNAR) approach to impute the time to event/censoring based on observed discontinuation and then present a combined estimate of treatment effect.

Missing data sensitivity analyses have become more and more essential to support robustness of primary conclusions, with regulatory agencies seeking for consistency with primary conclusions [11]. Overall, a lot of progress has been made in the past decade on continuous endpoints, with various methods based on different MNAR assumptions proposed. For instance, return to baseline [12, pp. 242–248, 13, pp. 641–653] assumes discontinued subjects will experience a washout effect and hence eventually return to their baseline levels. The class of control-based methods [14, pp. 1352–1371, 15, pp. 443–463] assumes subjects who discontinue the test treatment will follow the distribution of the reference group after discontinuation. Stress testing strategies such as tipping point analyses [16, pp. 1085–1098] is another very popular approach that has gained much attention in recent years. It searches for the breaking point under which statistical significance will vanish. These sensitivity analysis approaches can be adapted and readily applied to binary endpoints due to the connection between linear model and generalized linear model theory. Sensitivity analyses for time to event endpoints are, nevertheless, underdeveloped. There have been some key contributors in this area in the past decade. For instance, Jackson et al. [17, pp. 4681–4694] was the pioneer to propose a non-parametric bootstrap approach for imputation of missing data based on non-independent censoring (i.e., censoring not at random) and Ruau et al. implemented his proposed approach in an

R package [18]. Because Jackson's approach used a step function to estimate the cumulative hazard function  $H(t)$ , imputed time to event can only take the value of one of the observed times to event. As a result, his approach is not a full estimation approach and will not be applicable to a dataset with very few events due to the nature of the proposed procedures via step functions. Lipkovich et al. [19], pp. 216–229 treated missing data as non-ignorable and imputed them under tipping point framework (delta-adjusted), using piecewise exponential distribution and bootstrap sampling respectively. However, there have been no rules proposed in terms of which delta should be selected for a standalone sensitivity analysis. It's noteworthy that both Jackson and Lipkovich described the application of delta/gamma approach which serves as the foundation for tipping point sensitivity analysis. Zhao et al. [20, pp. 229–253] proposed linear interpolation to conduct imputation based on bootstrap sampling, using Kaplan–Meier (KM) and Cox proportional hazards model respectively. Because the field lacks a generalizable overall MNAR (or informative censoring) assumption that can be directly applied to missing data imputation, all above approaches use the full dataset as the imputation basis and heavily rely on a user-specified delta/gamma adjustment to achieve MAR or MNAR imputation.

In the era of estimands when more and more sponsors integrate estimands into the protocol framework, it dictates the need to align proposed sensitivity analyses with primary analyses on estimands, according to ICH E9 R1 [21]. In other words, after primary analysis and primary estimand are pre-specified in the protocol, the proposed sensitivity analyses must target the same estimand so that the results can be compared side by side. Because analyses of time to first events in CVOT are usually conducted under the intent to treat (ITT) principle with no exclusion of data post randomization, following the treatment policy (TP) strategy [22, pp. 1–19] there is no controversy that all subsequent sensitivity analyses should target the same TP estimand, in terms of handling of missing data that arise from intercurrent events.

It's not always easy to reach common ground between sponsors and regulators on the choice of sensitivity analyses. When sponsors propose a sensitivity analysis based on missing at random (MAR), regulatory agencies are interested in knowing if the primary conclusion will still hold after imposing some unfavorable assumed conditions. Therefore, it's critical and urgent to settle on something that's relatively acceptable to both parties. Assuming MAR for all discontinued subjects can be criticized, especially for subjects who discontinued due to reasons that might be plausibly related to their treatment (e.g., side effects), but imposing the worst assumptions on all missing data is also hard to justify, and ultimately, the choice of

assumptions for imputing missing data should be driven by the estimand of interest. The approach we propose can serve as a balance point between sponsor and regulatory agencies because its underlying MNAR assumption is quite neutral. The assumption states that subjects with missing data will have similar treatment effect compared to retrieved dropouts (defined as off-treatment subjects that remain in the study) in the same randomized group. It uses the subpopulation of “off-treatment” subjects to multiply impute the missing data.

Recently He et al. published a paper about retrieved dropout multiple imputation (MI) in CVOT using piecewise exponential distribution [22, pp. 1–19]. It was demonstrated using a lipid-lowering CVOT dataset, with respect to time to first MACE. Results of the piecewise exponential model were compared to the regular CPH model and the jump to reference multiple imputation. However, simulations were not conducted to inform type-I error and power rates. In our manuscript, we explore the approach in different ways: parametrically or non-parametrically with respect to the baseline hazard function. More specifically, we explore bootstrap, piecewise exponential and Weibull distributions for the imputation of time to event missing data. Simulation studies are conducted to help understand the type-I error and power rates under different clinical trial scenarios. Lastly, we illustrate these approaches by applying them to a real unblinded phase 3 CVOT.

The CVOT dataset needs to be broken down into 3 non-overlapping subsets, before the initiation of the analysis. Similar to the implementations in continuous endpoints [23, p. 82], retrieved dropouts (RDs) in outcome trials are usually defined as subjects who have discontinued randomized treatment but continue to stay in the trial. Because they are continuously followed until the occurrence of death or administrative censoring datacut, their outcome will be known. In the proposed analyses, they form the fundamental basis of missing data imputation, given that they resemble the missing data subjects in terms of treatment discontinuation and given the assumption that subjects who discontinued treatment will have similar outcomes whether they were followed up in the study or not. Subjects who have withdrawn consent or lost to follow-up will not take the assigned study medication, which is essentially identical to RDs except that RDs continue to stay in the study so that their outcome will be known. Completers, on the other hand, are defined as subjects who have completed the study while on study medication. Missing data is the remainder of the CVOT dataset, i.e., subjects who have discontinued from the study with no prior endpoint events. In the “Statistical Methods” section, we will first illustrate these methods with mathematical notation. Then we will summarize the implementation into an algorithm format. Next, we will conduct simulation

studies to understand and compare the performance of each approach under different clinical trial scenarios. The data analysis section provides a real application of these analyses to an unblinded CVOT dataset. Finally, all findings will be discussed and concluded.

## Statistical Methods

### The Basis of Time to Event Imputation

Assume in a CVOT  $2N$  subjects are randomized to two treatment groups in a 1:1 allocation ratio. Let  $T = T_i (i = 1, \dots, 2N)$  denote the time to first event/censoring of the primary endpoint. Then  $E_i$  is used to denote the event status: they are assigned a value of 1 for events and 0 otherwise.

We first model the conditional probability of a subject with missing data because this will enable the derivation of imputed time to first event. Let  $c$  denote the observed time to censoring, e.g. from randomization to study discontinuation,  $t$  is the unknown time to first event.  $S(t)$  stands for the survival function of  $T$ , then we have the following general equation:

$$P(T \leq t | T > c) = 1 - P(T > t | T > c) = 1 - \frac{S(t)}{S(c)} \in (0, 1). \quad (1)$$

$t$  needs to satisfy the condition that  $t > c$ , otherwise the above equation will reduce to 0. Then based on the above equation, we can randomly sample a variable from uniform distribution, i.e.  $u \sim \text{unif}(0, 1)$  and  $t$  can be estimated by solving the equation  $\frac{S(t)}{S(c)} = u$ . Due to the monotone property of  $S(t)$ , the estimated  $\hat{t}$  is ensured to be greater than  $c$ . Similarly this can be applied to all  $k$  subjects with missing data, by drawing  $k$  independent variables  $(u_1, \dots, u_k)$  from  $\text{unif}(0, 1)$ , s.t.

$$\frac{S_j(t_j)}{S_j(c_j)} = u_j (j = 1, \dots, k), \quad (2)$$

where  $S_j(c_j) = S(c_j | \text{tr}t_j, \mathbf{X}_j) = P(T > c_j | \text{tr}t_j, \mathbf{X}_j)$  are survival functions adjusting for treatment and baseline covariates and  $\hat{t}_j$  are the estimates of  $t_j$  by solving Eq. (2). For simplicity with no loss of generalization, let  $\text{tr}t_j = 1$  denote active treatment group and  $\text{tr}t_j = 0$  denote placebo group and  $\mathbf{X}_j$  is the vector of baseline covariates.

Next, we will describe how we combine this imputation step with the CPH model when the baseline hazard function  $h_0(t)$  is derived using: (1) non-parametric bootstrap sampling, (2) parametric distribution.

### Bootstrap

When an explicit parametric form is not imposed on baseline hazard function  $h_0(t)$ , just like in the semi-parametric Cox proportional hazards model, we propose to use bootstrap and interpolation to multiply impute the time to event conditional on observed time to censoring. Let  $C_j$  denote the individual time to study closure from randomization  $R_j$ . Without loss of generalization, assume the study uses a common study datacut  $D$  and  $C_j$  is defined as  $C_j = D - R_j + 1$  in days.

Assume there are a total of  $B$  bootstrap samples which generate  $B$  sets of survival function. Given a bootstrap sample  $b$ , the associated survival function is defined as  $S_j^{(b)}(t) = S^{(b)}(t|trt_j, \mathbf{X}_j) = P^{(b)}(T > t|trt_j, \mathbf{X}_j) (b = 1, \dots, B)$ . We first construct a grid with  $(m + 2)$  grid points as follows:  $c_j = t_j^{(0)} < t_j^{(1)} < t_j^{(2)} < \dots < t_j^{(m-1)} < t_j^{(m)} < t_j^{(m+1)} = C_j$ . Then  $S_j^{(b)}(c_j), S_j^{(b)}(t_j^{(1)}), \dots, S_j^{(b)}(C_j)$  can be derived using “predict” function along with “coxph” in R, written as  $\widehat{S}_j^{(b)}(c_j), \dots, \widehat{S}_j^{(b)}(C_j)$ . Since  $S_j^{(b)}(t_j)$  is derived as  $\widehat{S}_j^{(b)}(t_j) = S_j^{(b)}(c_j)u_j^{(b)}$  according to Eq. (2), we can identify interval, say  $(t_j^{(p)}, t_j^{(p+1)})$  in which  $t_j^{(b)}$  falls, s.t.  $\widehat{S}_j^{(b)}(t_j^{(p+1)}) \leq \widehat{S}_j^{(b)}(t_j) \leq \widehat{S}_j^{(b)}(t_j^{(p)})$ . Then  $t_j$  is estimated using linear interpolation, i.e.

$$\widehat{t}_j^{(b)} = t_j^{(p)} + \frac{t_j^{(p+1)} - t_j^{(p)}}{\widehat{S}_j^{(b)}(t_j^{(p+1)}) - \widehat{S}_j^{(b)}(t_j^{(p)})} \left( \widehat{S}_j^{(b)}(t_j) - \widehat{S}_j^{(b)}(t_j^{(p)}) \right). \tag{3}$$

If  $t_j^{(b)}$  falls outside the grids, then  $\widehat{t}_j^{(b)}$  is estimated as  $\widehat{t}_j^{(b)} = C_j$

### Parametric

Using parametric distribution to model  $h_0(t)$ , estimates of  $t_j$  can be derived in closed form. In the following sections, we demonstrate how to use piecewise exponential and Weibull distributions respectively to model the distribution of  $h_0(t)$  because they are able to represent distribution of most baseline hazard functions for time to event endpoints [24, pp. 682–701, 25, p. 873, 26, pp. 59–73, 27, pp. 10–11, 28, p. 152].

### Piecewise Exponential

When baseline hazard function follows a piecewise exponential distribution, the proportional hazards model is written as  $h(t|\mathbf{Z}_j; \boldsymbol{\theta}) = h_0(t)e^{\mathbf{Z}_j\boldsymbol{\beta}}$  where  $h_0(t) = \lambda_u$  for  $t$  in  $[\tau_{u-1}, \tau_u)$ , with  $0 = \tau_0 < \tau_1 < \dots < \tau_M = \infty$  as partitioning points to be estimated as well and  $\mathbf{Z}_j$  is the full design vector denoted as  $\mathbf{Z}_j = (trt_j, \mathbf{X}_j)$ . Its corresponding survival function is written as

$$S(t|\mathbf{Z}_j; \boldsymbol{\theta}) = \begin{cases} \exp(-\lambda_1 t e^{\mathbf{Z}_j\boldsymbol{\beta}}) & t \in [0, \tau_1) \\ \exp(-(\sum_{v=1}^{u-1} \lambda_v(\tau_v - \tau_{v-1}) + \lambda_u(t - \tau_{u-1}))e^{\mathbf{Z}_j\boldsymbol{\beta}}) & t \in [\tau_{u-1}, \tau_u) \\ \exp(-(\sum_{v=1}^{M-1} \lambda_v(\tau_v - \tau_{v-1}) + \lambda_M(t - \tau_{M-1}))e^{\mathbf{Z}_j\boldsymbol{\beta}}) & t \in [\tau_{M-1}, \infty) \end{cases} \tag{4}$$

A total of  $B$  imputations are planned. To make sure the imputation follows the “proper” principle [29, pp. 202–243, 30, pp. 473–489], different sets of coefficients and variance estimates have to be generated and applied to each imputation. The following two methods can fulfill this goal because they are asymptotically equivalent [31, pp. 593–607]. For SAS users, the most convenient way is to use “MCMC sampling” to draw Bayesian posterior samples for parameter coefficients  $\boldsymbol{\beta}$  and scale parameters  $\lambda_u (u = 1, \dots, M)$ . The alternative is to draw from a multivariate normal distribution with the MLE estimates as the mean and the estimated variance matrix of the MLE estimates as the variance (i.e., “MLE multivariate sampling”). Both approaches start from fitting a CPH model based on a piecewise exponential distribution using data of all RD subjects, with parameter estimates and estimate of the covariance matrix denoted as  $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}})$  and  $\widehat{var}(\widehat{\boldsymbol{\theta}})$ .

For each imputation ( $b = 1, \dots, B$ ), parameters  $\boldsymbol{\theta}^{(b)} = (\boldsymbol{\lambda}^{(b)}, \boldsymbol{\beta}^{(b)}) = (\lambda_1^{(b)}, \dots, \lambda_M^{(b)}, \boldsymbol{\beta}^{(b)})$  can be randomly sampled using either approach:

- *MCMC sampling*:  $\lambda_{m,(t+1)}^{(b)} \sim P(\lambda_m | \lambda_{1,(t+1)}^{(b)}, \dots, \lambda_{(m-1),(t+1)}^{(b)}, \lambda_{(m+1),(t)}^{(b)}, \dots, \lambda_{M,(t)}^{(b)}, \boldsymbol{\beta}^{(b)})$  for  $m = 1, \dots, M$  and continue through  $\boldsymbol{\beta}$  until iteration  $t$  is large enough and chain convergence is achieved.
- *MLE multivariate sampling*: sample  $\boldsymbol{\theta}^{(b)} \sim MVN(\widehat{\boldsymbol{\theta}}, \widehat{var}(\widehat{\boldsymbol{\theta}}))$ .

Then by combining the above survival function formula (4) with (2) and by letting  $S_j^{(b)}(t)$  be defined as  $S_j^{(b)}(t) = S(t|\mathbf{Z}_j; \boldsymbol{\theta}^{(b)})$ , the estimated time to first event is expressed as

$$\widehat{t}_j^{(b)} = \begin{cases} \tau_{M-1} - \frac{\log(u_{j*}^{(b)})e^{-\mathbf{Z}_j\boldsymbol{\beta}^{(b)}} + \sum_{v=1}^{M-1} \lambda_v^{(b)}(\tau_v - \tau_{v-1})}{\lambda_M^{(b)}} & u_{j*}^{(b)} = S_j^{(b)}(c_j)u_j^{(b)} \in (0, S_j^{(b)}(\tau_{M-1})] \\ \tau_{p-1} - \frac{\log(u_{j*}^{(b)})e^{-\mathbf{Z}_j\boldsymbol{\beta}^{(b)}} + \sum_{v=1}^{p-1} \lambda_v^{(b)}(\tau_v - \tau_{v-1})}{\lambda_p^{(b)}} & u_{j*}^{(b)} \in (S_j^{(b)}(\tau_p), S_j^{(b)}(\tau_{p-1})]; 2 \leq p < (M - 1) \\ \frac{-\log(u_{j*}^{(b)})e^{-\mathbf{Z}_j\boldsymbol{\beta}^{(b)}}}{\lambda_1^{(b)}} & u_{j*}^{(b)} \in (S_j^{(b)}(\tau_1), 1] \end{cases} \tag{5}$$

## Weibull

We first introduce accelerated failure time (AFT) [32, pp. 31–51, 33, pp. 1871–1879] model because the following AFT model is directly linked to Weibull regression. AFT model is very frequently used in real applications with statistical packages and software readily available [34, pp. 583–592].

$$\log T_j = \gamma_0 + \mathbf{Z}_j \boldsymbol{\gamma} + \sigma \varepsilon_j,$$

where  $\sigma$  is scale parameter and  $\varepsilon_j$  follows a Gumbel distribution (also known as extreme-value distribution).

Using the notations of Weibull AFT model, the hazard function based on Weibull distribution can be written as

$$h(t|\mathbf{Z}_j) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} e^{-(\mathbf{Z}_j \boldsymbol{\gamma} + \gamma_0)/\sigma} = h_0(t) e^{\mathbf{Z}_j \boldsymbol{\beta}} \text{ where}$$

$$\boldsymbol{\beta} = -\frac{\boldsymbol{\gamma}}{\sigma},$$

$$h_0(t) = \frac{1}{\sigma} \frac{t^{\frac{1}{\sigma}-1}}{e^{\frac{\gamma_0}{\sigma}}}. \quad (6)$$

Therefore, the baseline hazard function follows a Weibull distribution with  $\frac{1}{\sigma}$  as the shape parameter and  $e^{\gamma_0}$  as the scale parameter. Because the above AFT model is readily implemented in statistical software, e.g., “proc lifereg” in SAS software and “SurvRegCensCov” package in R, we will use notations of AFT model to derive imputed time to event as follows. To start with, we use  $\theta = (\sigma, \gamma_0, \boldsymbol{\gamma})$  to denote the vector of parameters.

For each imputation ( $b = 1, \dots, B$ ),  $\theta^{(b)}$  is sampled from either MCMC or MLE sampling. Then using  $S(t|\mathbf{Z}_j) = e^{-\exp(-\eta_j/\sigma)t^{\frac{1}{\sigma}}}$  with linear predictor  $\eta_j = \mathbf{Z}_j \boldsymbol{\gamma} + \gamma_0$  in conjunction with (2), time to first event is derived as

$$\hat{t}_j^{(b)} = (c_j \hat{\sigma}^{(b)} - \exp\left(\frac{\hat{\eta}_j^{(b)}}{\hat{\sigma}^{(b)}}\right) \log(u_j^{(b)}))^{\hat{\sigma}^{(b)}} \quad (7)$$

where  $\hat{\eta}_j^{(b)} = \mathbf{Z}_j \hat{\boldsymbol{\gamma}}^{(b)} + \hat{\gamma}_0^{(b)}$ .

## Algorithm

We summarize the general framework of implementation in this section.

A total of  $B$  multiple imputations/bootstrap samples are planned for imputation of missing data.

1. A Cox proportional hazards model is fit to the dataset comprised RDs (“RD dataset”) adjusting for treatment group and other pre-specified baseline covariates.

- *Bootstrap*:  $B$  different input datasets are created, by sampling from RD dataset with replacements. Then  $B$  different CPH models are fit, and each model will serve as the basis of the imputation.

- *Parametric*: One CPH model is fit, using the RD dataset. Then  $B$  different sets of parameters are sampled using either Bayesian MCMC sampling or MLE multivariate sampling (i.e., “proper imputation”).

For each imputation ( $b = 1, \dots, B$ ), proceed to step 2–4 sequentially.

2. Missing data of time to first event will be estimated using

- *Bootstrap*: Use formula (3) in conjunction with (2) with pre-specified grid points.
- *Parametric*: Use formula (5)/(7) in conjunction with (2).

3. Then imputed time to first event will be added to the randomization date to derive the imputed outcome.

- If the value is greater than the censoring data cut, then the imputed outcome will be deemed as a censored and his time to censoring will be derived as censoring data cut—randomization date + 1 in days.
- Otherwise, the imputed outcome will be deemed as an event and the imputed value will be used as the time to first event. *For studies that conduct vital search on subjects that are lost to follow up or have withdrawn consent, additional steps can be considered.* For instance, if final vital status is death, the imputed event date also needs to be compared to death date. More specifically,

- It will be imputed as an event if and only if the imputed event date + randomization date doesn’t exceed the earlier of death date and study data cut.
- Otherwise, it will be treated as censored and time to censoring will be derived as the earlier of death date and study data cut-randomization date + 1.

4. All imputed outcomes will be analyzed together with RDs and completers using the regular CPH model. Coefficient estimates of treatment effect and standard error (SE) will be generated.

5. Repeat step 2–4 for all  $B$  imputations. Coefficient estimates and SE are pooled and combined into a single estimate following Rubin’s rules [29, pp. 202–243].

## Simulation Studies

Simulation studies are conducted to evaluate the validity of the proposed approaches with respect to different clinical trial scenarios. Starting from type-I error simulations

we will first present and compare type-I error rate of the three proposed methods, under the null hypothesis that there is no difference of treatment effect between the two randomized groups, i.e.  $H_0 : HR = 1$  and with respect to different combinations of sample size, discontinuation rate and retrieved dropout rate, then we move to power simulations to evaluate how powerful these approaches are when the alternative hypothesis that the drug is superior to placebo is true, i.e.  $H_1 : HR < 1$ .

## Type-I Error

Five thousand datasets are simulated under the null hypothesis  $H_0 : HR = 1$ . Under each replicate, the trial is simulated as a two-arm randomized clinical trial with a 1:1 allocation ratio, assuming an overall event rate of 2% per person-year according to a few completed CVOT studies of 5–6 year duration [3, pp. 1713–1722, 5, pp. 1527–1539], a 1-year enrollment, a fixed duration of 5.5 years (i.e. a common censoring data cut of 5.5 years from the first randomization is applied). The following 27 scenarios with respect to different sample size, proportion of RDs and missing rate are explored:

- Sample size: 1000, 5000, 10,000;
- RDs: 5%, 20%, 40%;
- Overall discontinuation from the study, by EOS: 5%, 15%, 30%.

Time to event is simulated under exponential distribution. To test how robust the conclusions will hold, an additional distributional assumption is used. Time to discontinuation is simulated independently of events to avoid bias. RDs are randomly selected from the population set excluding study discontinuations to meet the definition and therefore there's no overlap between *RDs* and *Overall discontinuation from the study*. Empirical type-I error is defined as proportion of simulations significant at a significance level of two-sided 0.05. Results of type-I error are summarized in Table 2 and Supplementary Table 1.

## Power

A thousand datasets are simulated under the alternative hypothesis that the drug is superior to placebo with a 20% risk reduction, i.e.  $H_1 : HR = 0.8$ . The event rate in the placebo group is simulated as 2% per person-year, and the event rate in the active treatment group is modeled as  $2\% \times HR$  accordingly, with the exception that RDs in the active group is simulated as  $2\% \times HR^*$  ( $HR < HR^* < 1$ ) assuming RDs

in the active group will have a slightly reduced risk reduction due to off-treatment periods.  $HR^* = 0.9$  is used in this instance. The placebo discontinuation rate is simulated in the same way as in type-I error simulations. Whereas, the active group discontinuation rate is assumed to be lower by 2%, compared to placebo. Scenarios with sample size of 1000 is removed because a CVOT with 1000 subjects is very underpowered. The other parameters used in simulations remain the same as type-I error simulations.

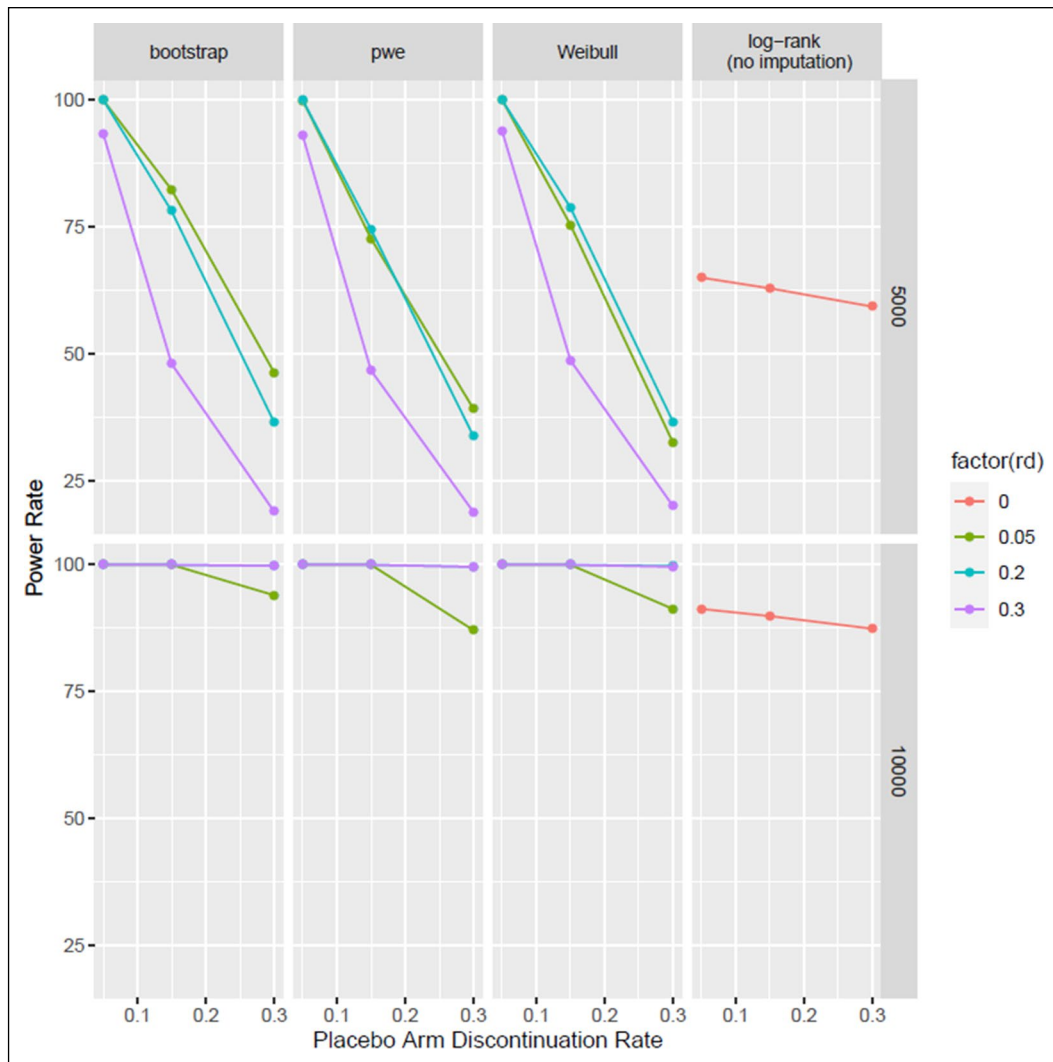
Empirical power rate is defined as proportion of simulations significant at a significance level of two-sided 0.05. In addition, power calculation based on the log-rank test is provided as a comparison reference which will help inform the conclusions. Results are summarized in Fig. 1.

## Data Analysis

We take a subset of a completed and unblinded phase III CVOT dataset to demonstrate applications of these 3 approaches. The subset is defined by restricting to subjects who have been randomized no later than year 4 (i.e., 4 years from the first randomization) and by shortening the study duration to 4 years (i.e., by applying a common censoring data cut of 4 years from the first randomization). In addition to the 3 approaches, we further explore jump to reference (j2r) imputation, and a regular CPH model and use them as comparison basis.

Within this snapshot there are a total of 7740 subjects (5143 on the active treatment group and 1998 on placebo) in this dataset, among which 1362 are RDs, 776 are premature discontinuations and 466 are considered as missing data. These characteristics are summarized in Table 1. The primary endpoint is time to first occurrence of 3-point MACE defined as a composite endpoint of non-fatal stroke, non-fatal MI and CV death. There are 599 first MACE events (379 in active vs 220 in placebo) without imputation. We summarize the  $n(\%)$  of events and censored prior to imputation by treatment group in Table 1. Missing data, RDs and completers are three mutually exclusive components of the dataset, adding up to 100%. For subjects in the category of “Have events prior to discontinuation”, if they are completely on-treatment leading to the first occurrence of MACE, they are considered completers; otherwise, they are considered RDs.

*J2r approach* assumes the survival function of the active group will follow the distribution of the placebo group (coded as 0 without loss of generality) following study discontinuation. In other words, formula (1) will be adapted to reflect the distribution switch for missing data in active treatment group (coded as 1), written as (where  $t > c$ )



**Figure 1** Compare the Power Rate of Three MIRD Approaches Under Different Clinical Scenarios

$$P((T \leq t | T > c) | trt = 1, X) = \frac{P(c < T \leq t | trt = 1, X)}{P(T > c | trt = 1, X)} = \frac{P(T > c | trt = 1, X) - P(T > t | trt = 0, X)}{P(T > c | trt = 1, X)} = 1 - \frac{S(t | trt = 0, X)}{S(c | trt = 1, X)}$$

Then the estimate  $\hat{t}$  will be derived by solving the equation  $\frac{S(t|trt=0,X)}{S(c|trt=1,X)} = u \sim U(0, 1)$ . In the multiple imputation setting  $\theta^{(b)}$  in  $\frac{S(t|trt=0,X;\theta^{(b)})}{S(c|trt=1,X;\theta^{(b)})} = u$  can be sampled using MCMC or MLE sampling using approximation or interpolation. In this case, we implement the J2R approach based on Weibull regression by means of MLE sampling because Weibull regression is considered as a substantive model [35, pp. 645–658]. The associated imputation formula is written as

$$\hat{t}_j^{(b)} = \max \left\{ \left( \exp \left( \frac{\hat{\eta}_{j,0}^{(b)} - \hat{\eta}_j^{(b)}}{\hat{\sigma}^{(b)}} \right) c_j^{\frac{1}{\hat{\sigma}^{(b)}}} - \exp \left( \frac{\hat{\eta}_{j,0}^{(b)}}{\hat{\sigma}^{(b)}} \right) \log(u_j^{(b)}) \right)^{\hat{\sigma}^{(b)}}, c_j \right\} \tag{9}$$

(8)

**Table 1** Disposition of Subjects in the CVOT Snapshot

	Active	Placebo
Randomized	5143	2598
Discontinued the study	501 (9.7%)	275 (10.6%)
<b>Missing data</b>	<b>306 (5.9%)</b>	<b>160 (6.2%)</b>
Have events prior to discontinuation	195 (3.8%)	115 (4.4%)
<b>RDs</b>	<b>835 (16.2%)</b>	<b>527 (20.3%)</b>
<b>Completers*</b>	<b>4002 (77.8%)</b>	<b>1911 (73.6%)</b>

\*Still in the study being followed by the snapshot date or have discontinued the study due to death

The 3 categories (Missing data, RDs and Completers) add up to the randomized population

**Table 2** Type-I Error Results for Different Scenarios, Evaluated at a Two-Sided Alpha = 0.05

N	Different Proportion of RDs	Proportion of Discontinuation <sup>#</sup>	Piecewise Exponential		Bootstrap		Weibull	
			Type-I Error	Avg Bias*	Type-I Error	Avg Bias*	Type-I Error	Avg Bias*
1000	0.05	0.05	0.0494	0.005	0.0518	-0.003	0.0352	-0.002
	0.2	0.05	0.0532	-0.001	0.0526	-0.003	0.0518	-0.003
	0.4	0.05	0.0528	-0.002	0.0524	-0.003	0.0524	0.002
	0.05	0.15	0.0386	0.018	0.0568	-0.003	0.0128	-0.005
	0.2	0.15	0.0532	0.003	0.0540	-0.004	0.0506	-0.004
	0.4	0.15	0.054	-0.001	0.0550	-0.004	0.0532	-0.004
	0.05	0.3	0.0254	0.037	0.0526	-0.005	0.0567	-0.0002
	0.2	0.3	0.0529	0.010	0.0542	-0.003	0.0464	-0.003
5000	0.4	0.3	0.0551	0.003	0.0554	-0.004	0.0534	-0.004
	0.05	0.05	0.0538	0.002	0.0528	<0.001	0.0526	<0.001
	0.2	0.05	0.0514	0.001	0.0510	<0.001	0.0506	<0.001
	0.4	0.05	0.0526	<0.001	0.0526	<0.001	0.0528	<0.001
	0.05	0.15	0.0488	0.005	0.0528	<0.001	0.047	<0.001
	0.2	0.15	0.0514	0.002	0.0514	<0.001	0.0492	<0.001
	0.4	0.15	0.0534	0.001	0.0520	<0.001	0.0518	<0.001
	0.05	0.3	0.0420	0.010	0.0492	-0.0002	0.0356	<0.001
10,000	0.2	0.3	0.0516	0.002	0.0500	-0.0003	0.0500	-0.0002
	0.4	0.3	0.0514	0.001	0.0523	-0.0004	0.0516	<0.001
	0.05	0.05	0.0536	0.002	0.0524	0.0015	0.0526	0.0015
	0.2	0.05	0.0516	0.002	0.0514	0.0015	0.0508	0.0015
	0.4	0.05	0.0510	0.002	0.0525	0.001	0.0514	0.0015
	0.05	0.15	0.0484	0.005	0.0482	0.002	0.0468	0.0020
	0.2	0.15	0.0506	0.002	0.0485	0.001	0.0490	0.0019
	0.4	0.15	0.0510	0.002	0.0510	0.0026	0.0504	0.0019
	0.05	0.3	0.0456	0.007	0.0470	0.0019	0.0424	0.0025
	0.2	0.3	0.0488	0.004	0.0478	0.0029	0.0496	0.0025
	0.4	0.3	0.0488	0.003	0.0470	0.0019	0.0492	0.0025

<sup>#</sup>Missing rate is lower than proportion of study discontinuation because subjects with events prior to study discontinuation are not counted as missing data

\*On log(HR) scale

## Results

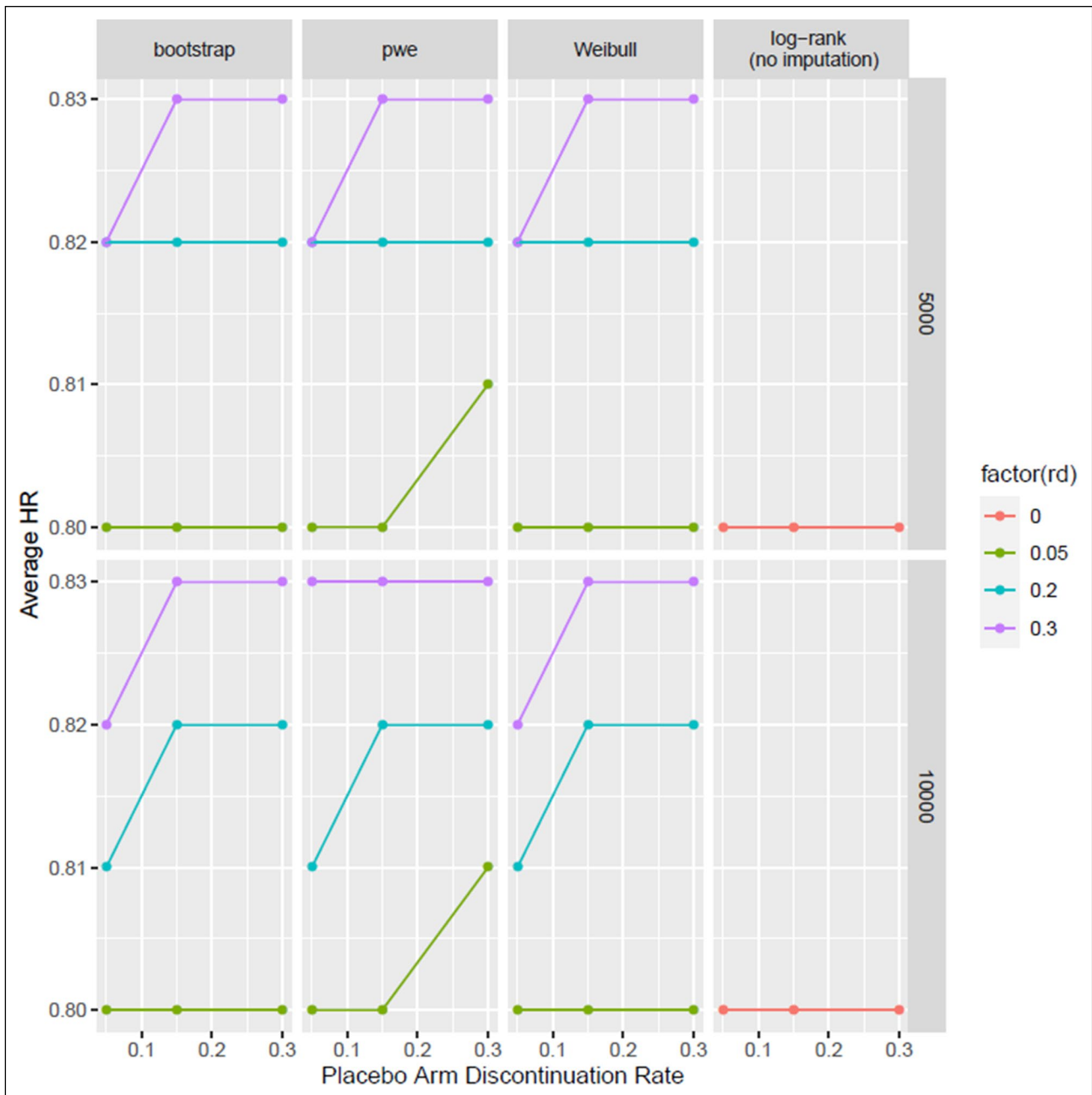
### Type-I Error

As shown in Table 2, all three approaches well control the type-I error rate and the bias relative to 0 is negligible (in log scale, 0 corresponding to HR = 1). It's noteworthy in small sample size scenarios where the RDs and missing data are very imbalanced, the type-I error of piecewise exponential distribution seems a little deflated. For instance, when there are 5% RDs and 30% missing data in a study with a total of 1000 subjects, piecewise exponential distribution returns slightly deflated type-I error rate while the type-I error rate of the other two is close to 0.05 (Table 2).

### Power Rate

All three approaches have very similar power rate, with the bootstrap approach slightly more powerful than the other two (Fig. 1). Furthermore, all three approaches follow similar patterns: (1) given a sample size and an RD rate, more missing data will lead to some power loss; (2) When sample size is large and given a study discontinuation rate, higher proportion of RDs results in a slightly higher power compared to studies with a lower proportion. (3) When sample size is small to medium, given a study discontinuation rate, low RD rates lead to higher power rates. It's very noteworthy in some scenarios the proposed approaches are more powerful than log-rank test assuming no RDs in the study. Different competing forces that contribute to these phenomena will be discussed in the "Discussion" section.





**Figure 2** Compare the HR of Three MIRD Approaches Under Different Clinical Scenarios

From the perspective of hazard ratio, all three approaches lead to slightly reduced risk reduction (i.e. larger HR) after adjusting for diluted treatment effect contributed by RDs (Fig. 2). Within the same sub plot, more RDs lead to a more reduced risk reduction (i.e., a larger HR); for instance, the HR corresponding to 20% RDs is larger than the one with 5% RDs.

### Data Analysis

All imputation approaches lead to an increase in number of events, compared to the regular CPH. Among all approaches, J2R leads to the highest increase in number of events, especially in the active group which explains why its estimated risk reduction ( $= 1 - HR$ ) is the smallest of all. All three approaches yield very similar results consistently, with risk reduction all larger than J2R but slightly smaller than CPH. For the bootstrap version,

**Table 3** Results of Using Different Statistical Methods to Analyze Time to First MACE

Method	Events in placebo	Events in Active	HR (95% CI) vs Placebo	p-Value
cph	220	379	0.86 (0.73, 1.02)	0.08
j2r	224.2	392.8	0.87 (0.74, 1.03)	0.11
MIRD_weibull	222.5	385.8	0.86 (0.73, 1.02)	0.09
MIRD_pwe*	222.4	386.9	0.87 (0.73, 1.02)	0.09
MIRD_bootstrap (grid width = 60 days)	222.5	385.4	0.86 (0.73, 1.02)	0.08
MIRD_bootstrap (grid width = 90 days)	222.4	386.9	0.87 (0.73, 1.02)	0.09

\*4 piecewise exponential hazards are used with cut points generated using  $\text{cuts} < -c(0, \text{quantile}(A\$t2e) [2:4], \text{max}(A\$t2e) + 1)$  before passed to the piecewise function

using more refined grids will lead to an estimated risk reduction very similar to the CPH, according to Table 3.

## Software Implementation

SAS is recommended for implementation of MCMC sampling because Bayesian sampling is readily embedded in several relevant SAS procedures such as Proc Lifereg, Proc MCMC. In combination with these SAS procedures, it's adequate to use SAS for imputation and data analyses. For example, users can firstly use Proc lifereg to generate the Bayesian sampling of the model parameters, and then perform matrix manipulation through Proc IML for time to events imputation, followed by Proc Phreg for the final analysis by imputation. Nevertheless, overall, it still poses some computing challenges for simulations, especially for approaches involving intensive interpolations and sampling. R is used in all simulations of this manuscript with the use of MLE sampling, due to the asymptotic equivalence [36, pp. 473–483]. The only limitation exists in piecewise exponential imputation because the current R package “eha” only provides a partial covariance matrix estimate, restricting to regression coefficients only (i.e. without piecewise baseline hazards). All R functions and SAS macros are provided in the supplementary files for reference.

## Discussion

Cox proportional hazards model has long been regarded as the gold standard to analyze time to first event endpoints in cardiovascular and renal outcome trials [37, pp. 1425–1435, 38, pp. 896–907, 39, pp. 1436–1446, 40, pp. 841–851, 41, pp. 2295–2306, 5, pp. 1527–1539, 42, pp. 347–357, 43, pp. 2117–2128]. On one hand, it's still frequently used as primary analysis as a well-established model. On the other hand, it has been noted that censoring for premature study discontinuation, known as presumed “non-informative censoring” can result in potential bias in the

treatment effect estimate across different therapeutic areas [44, pp. 2001–2009, 10, p. 40, 45, p. 101865, 9, p. 4, 46, pp. 327–328, 47, pp. 1433–1440]. In our proposed analyses, subjects censored due to non-administrative reasons are treated as missing data if they don't have relevant events prior to the discontinuation, and the potential for informative censoring is allowed by imputing their outcome assuming they follow the distribution of RDs in the same randomized group. Generally, they are not considered as missing due to random reasons, and we aim to impute their outcome after their study discontinuation, based on the assumption that their time to event distribution will approximately follow the distribution of RDs in the same randomized group. As a MNAR assumption, this serves as the foundation for implementation of the imputation. The subsequent analysis and combining steps are common to almost all MI approaches.

Multiple imputation is favored over single imputation because it allows uncertainty by means of multiple sampling, regardless of whether the sampling is parametric or not. Furthermore, it leads to non-biased estimates by using Rubin's rule which combines within and between imputation variability [29, pp. 202–243]. Unlike the continuous MI-RD approach, implementation of time to event MI-RD approach doesn't have small sample size issues in CVOT. Because size of outcome trials is normally large and usually it has more RDs than missing data, there is a high degree of plausibility that the model-based standard error is sufficiently small to get imputation values that closely approximate the results that would have been obtained in the real world. Generally speaking, 5–20% RDs in a CVOT will enable the implementation of the MI-RD approach.

All three variations have well-controlled type-I error rate. In terms of power rate, there are scenarios under which these approaches lead to improved power rate, as a result of interactions of two “competing” forces. The first one refers to increase in number of events from imputation, while the other one refers to attenuated risk reduction (i.e., larger HR) after imputation. More specifically, these MI approaches based on RDs tend to increase the overall hazard rate of the active group post-imputation due to a slightly increased hazard rate in the active group during the off-treatment period. Time to event endpoints are mainly powered by total number

of events and the alternative HR assumption. The first one has the potential to increase power via the additional events added by imputation while the latter has the tendency to reduce power via attenuation of treatment effect toward an HR that is closer to 1.0. The scenarios in which the first force outweigh the 2nd one are those with power gain over traditional log-rank tests assuming no RDs (Fig. 1). In the setting of a typical superiority CVOT with 4000 to 10,000 randomized subjects, 15–25% RD rate and 1% annualized study discontinuation rate, our proposed approaches will lead to increased power rates. For smaller CVOTs (first two rows in Fig. 1), more RDs lead to reduced power rates in every sub plot because the impact of treatment effect attenuation outweighs increase in events post-imputation.

The proposed three approaches can be implemented and applied to real clinical datasets in either R or SAS on PC platform. They have very similar performance but slightly different computational complexity. For each approach, users can either choose MCMC sampling or MLE sampling because they are asymptotically equivalent. Although the current R implementation for piecewise exponential distribution using R package “eha” only allows partial sampling of parameters, the type-I error and power rate assessments do align with the results of other approaches. Users can find all R functions and SAS examples in the supplementary files. No matter which approach the sponsor pre-specifies for the clinical study, it’s essential to gain consensus with the health authorities (HA) before unblinding a study.

In this manuscript, we propose an example of using a MNAR assumption to impute missing data of time to event endpoints. There are various ways to pre-specify and to perform sensitivity analyses. The fundamentals are to decide which assumption best fits the purpose. A conservative assumption has the potential to solidify the primary conclusion if the results don’t alter the primary conclusion, but it also has the risk of deviating from primary results. Because our assumption is neither conservative nor aggressive, it should be relatively acceptable to both HA and sponsors.

## Conclusion

We propose three MI approaches based on the same MNAR assumption for time to event endpoints in CVOT. They truly are estimation approaches that can be implemented using non-parametric bootstrap or parametric methods via either MCMC or MLE multivariate sampling (“proper imputations”) due to asymptotic equivalence. These approaches can be readily extended to studies in other therapeutic areas if the trials continue to follow patients regardless of treatment discontinuation or not. The three proposed approaches have very similar type-I error and power rates given a clinical scenario, with bootstrap being the most optimal solution due

to the nature of the approach that it takes into account uncertainty by not imposing parametric restriction. It’s noteworthy that bootstrap approach is more computationally intensive, while Weibull regression is the least. When bootstrap computation is a burden, Weibull or piecewise regression will suffice according to the supplementary analysis conducted: These two approaches are also robust when the underlying data deviates from the parametric distribution. Because the underlying assumption is not as conservative as jump to reference multiple imputation, the estimated HR usually is smaller than jump to reference analyses but larger than Cox model. We believe the proposed MI approaches meet the expectation of health authorities in terms of their capabilities to justify robustness of primary conclusions. Furthermore, they can serve as primary analysis and driver of sample size estimation.

## Acknowledgements

The authors would like to thank the two anonymous reviewers for carefully reviewing and providing very insightful comments that have led to further improvements.

## Author Contributions

SW contributed to the conception of the manuscript. SW drafted the manuscript. SW, RF, JPM revised, reviewed, and approved the manuscript.

## Funding

There’s no additional financial support beyond the authors’ employment.

## Data availability

**To protect the patients’ privacy, the real CVOT dataset will not be accessible to the public.**

## Declarations

## Conflict of interest

The authors have no conflict of interest.

## Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s43441-023-00575-5>.

## References

- FDA Center for Drug Evaluation and Research. *Guidance for industry: diabetes mellitus—evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes*. Beltsville: FDA Center for Drug Evaluation and Research; 2008.
- Food and Drug Administration Center for Drug Evaluation and Research (CDER). *Type 2 Diabetes Mellitus: Evaluating the Safety of New Drugs for Improving Glycemic Control Guidance for Industry*. 2020. <https://www.fda.gov/media/135936/download>.
- Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, Kuder JF, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med*. 2017;376(18):1713–22. <https://doi.org/10.1056/NEJMoa1615664>.
- Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, Darius H, et al. Ezetimibe added to statin therapy after acute coronary syndromes. *N Engl J Med*. 2015;372(25):2387–97. <https://doi.org/10.1056/NEJMoa1410489>.
- Ridker PM, Revkin J, Amarenco P, Brunell R, Curto M, Civeira F, Flather M, et al. Cardiovascular efficacy and safety of bococizumab in high-risk patients. *N Engl J Med*. 2017;376(16):1527–39. <https://doi.org/10.1056/NEJMoa1701488>.
- Lagakos SW. General right censoring and its impact on the analysis of survival data. *Biometrics*. 1979;35(1):139–56. <https://doi.org/10.2307/2529941>.
- Andersen PK. Censored data. In: Kotz S, Read CB, Balakrishnan N, Vidakovic B, Johnson NL, editors. *Encyclopedia of statistical sciences*. Hoboken: Wiley; 2013. p. 1–6.
- Kleinbaum DG, Klein M. *Survival analysis a self-learning text*. Berlin: Springer; 2012.
- Shih WJ. Problems in dealing with missing data and informative censoring in clinical trials. *Curr Control Trials Cardiovasc Med*. 2002;3(1):4. <https://doi.org/10.1186/1468-6708-3-4>.
- Ranganathan P, Pramesh C. Censoring in survival analysis: potential for bias. *Perspect Clin Res*. 2012;3(1):40. <https://doi.org/10.4103/2229-3485.92307>.
- European Medicines Agency. *Guideline on missing data in confirmatory clinical trials*. Amsterdam: European Medicines Agency; 2010.
- Zhang Y, Zimmer Z, Xu L, Lam RLH, Huyck S, Golm G. Missing data imputation with baseline information in longitudinal clinical trials. *Stat Biopharm Res*. 2022;14(2):242–8. <https://doi.org/10.1080/19466315.2020.1845234>.
- Qu Y, Dai B. Return-to-baseline multiple imputation for missing values in clinical trials. *Pharm Stat*. 2022;21(3):641–53. <https://doi.org/10.1002/pst.2191>.
- Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J Biopharm Stat*. 2013;23(6):1352–71. <https://doi.org/10.1080/10543406.2013.834911>.
- Cro S, Morris TP, Kenward MG, Carpenter JR. Reference-based sensitivity analysis via multiple imputation for longitudinal trials with protocol deviation. *Stand Genom Sci*. 2016;16(2):443–63. <https://doi.org/10.1177/1536867X1601600211>.
- Yan X, Lee S, Li N. Missing data handling methods in medical device clinical trials. *J Biopharm Stat*. 2009;19(6):1085–98. <https://doi.org/10.1080/10543400903243009>.
- Jackson D, White IR, Seaman S, Evans H, Baisley K, Carpenter J. Relaxing the independent censoring assumption in the cox proportional hazards model using multiple imputation. *Stat Med*. 2014;33(27):4681–94. <https://doi.org/10.1002/sim.6274>.
- Ruau D. *Gamma Imputation Tutorial (Jackson 2014)*. 2023.
- Lipkovich I, Ratitch B, O’Kelly M. Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. *Pharm Stat*. 2016;15(3):216–29. <https://doi.org/10.1002/pst.1738>.
- Zhao Y, Herring AH, Zhou H, Ali MW, Koch GG. A multiple imputation method for sensitivity analyses of time-to-event data with possibly informative censoring. *J Biopharm Stat*. 2014;24(2):229–53. <https://doi.org/10.1080/10543406.2013.860769>.
- U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials Guidance for Industry*. 2021.
- He J, Crackel R, Koh W, Chen L-W, Li F, Zhang J, Rothmann M. Retrieved-dropout-based multiple imputation for time-to-event data in cardiovascular outcome trials. *J Biopharm Stat*. 2022;33:234. <https://doi.org/10.1080/10543406.2022.2118763>.
- Wang S, Hu H. Impute the missing data using retrieved dropouts. *BMC Med Res Methodol*. 2022;22(1):82. <https://doi.org/10.1186/s12874-022-01509-9>.
- Carroll KJ. On the use and utility of the weibull model in the analysis of survival data. *Control Clin Trials*. 2003;24(6):682–701. [https://doi.org/10.1016/S0197-2456\(03\)00072-2](https://doi.org/10.1016/S0197-2456(03)00072-2).
- Plana D, Fell G, Alexander BM, Palmer AC, Sorger PK. Cancer patient survival can be parametrized to improve trial precision and reveal time-dependent therapeutic effects. *Nat Commun*. 2022;13(1):873. <https://doi.org/10.1038/s41467-022-28410-9>.
- Han G, Schell MJ, Kim J. Improved survival modeling in cancer research using a reduced piecewise exponential approach. *Stat Med*. 2014;33(1):59–73. <https://doi.org/10.1002/sim.5915>.
- Han G, Zhao B, Pye K, Zhao H. The piecewise exponential distribution. *Significance*. 2017;14(6):10–1. <https://doi.org/10.1111/j.1740-9713.2017.01086.x>.
- Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013;13(1):152. <https://doi.org/10.1186/1471-2288-13-152>.
- Rubin DB. Procedures with nonignorable nonresponse. In: Rubin DB, editor. *Multiple imputation for nonresponse in surveys*. Berlin: Wiley; 1987. p. 202–43.
- Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91(434):473–89. <https://doi.org/10.2307/2291635>.
- Nielsen SF. Proper and improper multiple imputation. *Int Stat Rev*. 2003;71(3):593–607.
- Kalbfleisch J, Prentice R. Failure time models. In: Kalbfleisch JD, Prentice RL, editors. *The statistical analysis of failure time data*. Hoboken: Wiley; 2002. p. 31–51.
- Wei LJ. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Stat Med*. 1992;11(14–15):1871–9. <https://doi.org/10.1002/sim.4780111409>.
- Odell PM, Anderson KM, Kannel WB. New models for predicting cardiovascular events. *J Clin Epidemiol*. 1994;47(6):583–92. [https://doi.org/10.1016/0895-4356\(94\)90206-2](https://doi.org/10.1016/0895-4356(94)90206-2).

35. Atkinson A, Kenward MG, Clayton T, Carpenter JR. Reference-based sensitivity analysis for time-to-event data. *Pharm Stat*. 2019;18(6):645–58. <https://doi.org/10.1002/pst.1954>.
36. Geyer CJ. Practical Markov chain Monte Carlo. *Stat Sci*. 1992;7(4):473–83.
37. Cannon CP, Pratley R, Dagogo-Jack S, Mancuso J, Huyck S, Masiukiewicz U, Charbonnel B, et al. Cardiovascular outcomes with ertugliflozin in type 2 diabetes. *N Engl J Med*. 2020;383(15):1425–35. <https://doi.org/10.1056/NEJMoa2004967>.
38. Gerstein HC, Sattar N, Rosenstock J, Ramasundarahettige C, Pratley R, Lopes RD, Lam CSP, et al. Cardiovascular and renal outcomes with efpeglenatide in type 2 diabetes. *N Engl J Med*. 2021;385(10):896–907. <https://doi.org/10.1056/NEJMoa2108269>.
39. Heerspink HJL, Stefánsson BV, Correa-Rotter R, Chertow GM, Greene T, Hou F-F, Mann JFE, et al. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med*. 2020;383(15):1436–46. <https://doi.org/10.1056/NEJMoa2024816>.
40. Husain M, Birkenfeld AL, Donsmark M, Dungan K, Eliaschewitz FG, Franco DR, Jeppesen OK, et al. Oral semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2019;381(9):841–51. <https://doi.org/10.1056/NEJMoa1901118>.
41. Perkovic V, Jardine MJ, Neal B, Bompoint S, Heerspink HJL, Charytan DM, Edwards R, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med*. 2019;380(24):2295–306. <https://doi.org/10.1056/NEJMoa1811744>.
42. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Silverman MG, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2019;380(4):347–57. <https://doi.org/10.1056/NEJMoa1812389>.
43. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*. 2015;373(22):2117–28. <https://doi.org/10.1056/NEJMoa1504720>.
44. López-López JA, Sterne JAC, Higgins JPT. Selection bias introduced by informative censoring in studies examining effects of vaccination in infancy. *Int J Epidemiol*. 2019;48(6):2001–9. <https://doi.org/10.1093/ije/dyz092>.
45. Riestler K, Kappos L, Selmaj K, Lindborg S, Lipkovich I, Elkins J. Impact of informative censoring on the treatment effect estimate of disability worsening in multiple sclerosis clinical trials. *Mult Scler Relat Disord*. 2020;39:101865. <https://doi.org/10.1016/j.msard.2019.101865>.
46. Templeton AJ, Amir E, Tannock IF. Informative censoring—a neglected cause of bias in oncology trials. *Nat Rev Clin Oncol*. 2020;17(6):327–8. <https://doi.org/10.1038/s41571-020-0368-0>.
47. Wilson BE, Nadler MB, Desnoyers A, Amir E. Quantifying withdrawal of consent, loss to follow-up, early drug discontinuation, and censoring in oncology trials. *J Natl Compr Cancer Netw*. 2021;19(12):1433–40. <https://doi.org/10.6004/jnccn.2021.7015>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.