

Distributional imputation for the analysis of censored recurrent events

Sarah R. Fairfax¹ | Shu Yang¹

Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA

Correspondence

Sarah R. Fairfax, Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA.

Email: seriegel@ncsu.edu

Funding information

National Institutes of Health, Grant/Award Numbers: 1R01AG066883, 1R01ES031651; National Science Foundation, Grant/Award Number: SES 2242776

Longitudinal clinical trials for which recurrent events endpoints are of interest are commonly subject to missing event data. Primary analyses in such trials are often performed assuming events are missing at random, and sensitivity analyses are necessary to assess robustness of primary analysis conclusions to missing data assumptions. Control-based imputation is an attractive approach in superiority trials for imposing conservative assumptions on how data may be missing not at random. A popular approach to implementing control-based assumptions for recurrent events is multiple imputation (MI), but Rubin's variance estimator is often biased for the true sampling variability of the point estimator in the control-based setting. We propose distributional imputation (DI) with corresponding wild bootstrap variance estimation procedure for control-based sensitivity analyses of recurrent events. We apply control-based DI to a type I diabetes trial. In the application and simulation studies, DI produced more reasonable standard error estimates than MI with Rubin's combining rules in control-based sensitivity analyses of recurrent events.

KEY WORDS

control-based imputation, distributional imputation, intercurrent events, recurrent events, sensitivity analysis

1 | INTRODUCTION

Recurrent events, such as recurrences of asthma exacerbations or tumor regrowths, are pertinent in many clinical areas and arise when subjects are at risk of experiencing multiple incidences of the same event. Recurrent event endpoints are prevalent particularly in chronic disease areas. For example, in a clinical trial conducted by the Juvenile Diabetes Research Foundation (JDRF) Continuous Glucose Monitoring (CGM) Research Group to evaluate the effect of CGM on the management of type I diabetes, incidences of severe hypoglycemic events were collected as a recurrent events outcome. A typical question of interest in such trials with recurrent event endpoints is whether a proposed treatment affects the number or rate of events during the time of planned follow-up. In the context of the JDRF trial, for instance, researchers may wish to assess whether CGM decreases the expected number or rate of hypoglycemic events for patients with type I diabetes. Standard estimands to quantify the treatment effect in recurrent events trials then include the change in expected rate of events, the rate ratio of events, or the rate of the instantaneous probability of having an event when comparing active treatment to control. Motivated by the clinical question of interest, recurrent events may be modeled by either event counts or time to events.

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

© 2024 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

In longitudinal clinical trials, particularly when planned follow-up time is long or many adverse effects are expected, missing event data are unavoidable. Missing data can arise for a variety of reasons, such as loss to follow-up or deviations from protocol, which could be due to administrative reasons or reasons related to treatment assignment itself. There are three common assumptions for the missing data mechanism: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).¹ Data are said to be MCAR if the pattern of missingness depends neither on observed nor missing data values. For the missing mechanism to be MAR, the missingness must only be dependent on values observed during the study. If, conversely, the missing data process is dependent on the missing values or other unknown measures, the missingness is MNAR. In this paper, we will primarily consider missingness due to dropout and loss to follow-up.

The primary analysis in clinical trials is typically performed assuming data are MAR,² but this assumption is often not verifiable in practice. The ICH E9(R1) addendum urges that intercurrent events leading to missing data be accounted for in defining treatment effect estimands, and recommends sensitivity analyses to assess the robustness of study conclusions to unverifiable assumptions in the primary analysis.³ In the language of the ICH E9(R1) addendum, control-based imputation⁴ can be an effective “hypothetical strategy,” wherein the event rate or intensity ratio is estimated under hypothetical assumptions about how and why data are missing post-dropout.⁵ Control-based imputation involves assuming subjects with missing data on the active treatment arm have event profiles more similar to those observed on the control arm. This assumption is reasonable particularly when subjects are not expected to receive treatment post-dropout or when the reference arm is the standard of care treatment. For example, in the JDRF CGM type I diabetes trial, subjects were randomized to either standard self-monitoring of blood glucose alone or CGM in combination with self-monitoring blood glucose. If participants randomized to the CGM arm withdrew from the study, they would be likely to continue self-monitoring blood glucose, making the assumption that their post-dropout behavior would be similar to that observed for control subjects more plausible.

Control-based imputation targets the *de facto* estimand,⁴ which assesses what the effect of treatment would be in practice rather than in the ideal situation where every subject adheres completely to the protocol. Control-based imputation is also an attractive option for sensitivity analyses in superiority trials from a regulatory standpoint, as it tends to yield more conservative treatment effect estimands.⁶ A practical benefit of control-based imputation is that it does not require identification of distributions of unknown parameters to describe the behavior of missing events due to its foundation in pattern-mixture models,⁷ which are instead built under specification of the patterns of behavior of missing events given the observed data.² This pattern-mixture model framework encourages clear, clinically meaningful assumptions about differences in patterns between subjects who complete the study and those who drop out.⁸

Multiple imputation (MI)⁹ is a popular approach to applying control-based assumptions, but Rubin’s standard combining rules for variance estimation have been shown to be inaccurate for estimation of the true sampling variability of MI treatment effect estimators in the control-based setting.¹⁰⁻¹² A variety of analytical variance estimators^{6,10,11,13-15} have been proposed for improvements over Rubin’s combining rules, but these estimators often involve complicated, model-specific formulas that make their implementation with recurrent events challenging. A more common approach to improving variance estimation with control-based MI of recurrent events is the nonparametric bootstrap,^{5,16,17} but this approach can be very computationally intensive. In light of the limitations of these existing variance estimation methods in applications of MI to recurrent events, we propose distributional imputation (DI) as a promising alternative to MI for control-based sensitivity analyses of recurrent events endpoints. Like MI, DI offers procedural simplicity for implementing control-based assumptions. We also introduce a complementary wild bootstrap procedure for estimation of treatment effect estimator variance under DI that is more flexible than analytical variance estimators and more efficient than the nonparametric bootstrap.

In this article, we first introduce the notation and models to be used in implementing our method in the recurrent events framework and ground our discussion in the context of a real-world type I diabetes trial. We next review a few control-based assumptions commonly applied in sensitivity analyses of recurrent events endpoints, as well as MI, the conventional method of applying these control based assumptions. Next, we introduce DI and the accompanying wild bootstrap variance estimation procedure, and detail the asymptotic behavior and results of the DI point and variance estimators. We then evaluate the finite sample performance of DI compared to MI with either Rubin’s combining rules or a nonparametric bootstrap for the control-based imputation of recurrent events. Finally, we return to our type I diabetes trial example, for which we compare the execution of DI to that of MI with Rubin’s combining rules and a Bayesian approach to MI with the nonparametric bootstrap. We conclude with a discussion of our method and future work.

2 | NOTATION AND MODEL SPECIFICATION

2.1 | Basic setup

We first introduce the notation and models we will utilize in the application of DI for recurrent events. The two main approaches to modeling recurrent events include modeling event counts and modeling time to recurrent events by gap or total time. Count models are the simpler of these approaches, but we will incorporate both methods within the context of our recurrent event sensitivity analysis setting.

We consider a 2-arm randomized trial where A_i indicates whether subject i is randomized to active treatment ($A_i = 1$) or control ($A_i = 0$). Suppose we have a nonhomogeneous Poisson counting process, $N_i(t)$, for subjects $i \in \{1, \dots, n\}$. The length of planned follow-up time for subject i is τ_i , which may vary among subjects. For example, τ_i would vary among subjects if enrollment is done on a rolling basis with follow-up planned to continue until the set end of the study. Conversely, if follow-up is planned for a set amount of time from randomization regardless of enrollment date, then τ_i would be constant for all i .

Suppose $N_i(t)$ is subject to possible censoring due to dropout, and subjects are followed until time $C_i \leq \tau_i$. Then the full outcome, $N_i(\tau_i)$, for censored subjects may be deconstructed into observed and missing portions, $N_i(\tau_i) = \{N_i(C_i), N_i(\tau_i - C_i)\}$, respectively, where the event rate is assumed constant for each subject i . For simplicity, we assume missingness is monotone in that subjects who withdraw are not re-entered into the study. Let R_i be an indicator of censoring such that $R_i = 1$ if subject i is followed to completion and $R_i = 0$ if $C_i < \tau_i$ for subject i . Suppose $N_i(t)$ is also dependent on possible baseline covariates, Z_i . Call $X_i^T = (A_i, Z_i^T)$ and $\zeta^T = (\delta, \beta)$.

We assume $N_i(t)$, $i \in \{1, \dots, n\}$, follows a proportional intensity model with gamma frailty,

$$\Lambda_{X_i}(t, b_i, \theta) = \Lambda_0(t, \lambda)b_i \exp(A_i\delta + Z_i^T\beta), \quad (1)$$

where we assume the baseline cumulative intensity $\Lambda_0(t, \lambda)$ is parametric and b_i are i.i.d. from a gamma distribution with mean 1 and variance γ . The full model parameter is then $\theta^T = (\lambda, \delta, \beta, \gamma)$. This frailty model leverages the time to events in modeling the event intensity, and allows the event intensity to vary by subject-specific random effect, b_i , representing the excess risk of events for subject $i \in \{1, \dots, n\}$ due to unknown factors.¹⁸ We also use the Anderson-Gill formulation,¹⁹ $\exp(X_i^T\zeta)$, for the portion of $\Lambda_{X_i}(t, b_i, \theta)$ dependent on X_i to avoid restrictions on δ and β .

Given this nonhomogeneous Poisson process, let Y_i be the event count for subject i during planned follow-up, τ_i . For subjects followed until time $C_i < \tau_i$, the full event count can be written $Y_i = (Y_{i,obs}, Y_{i,mis})$, where $Y_{i,obs}$ is the observed event count during $(0, C_i]$ and $Y_{i,mis}$ is the unobserved event count during $(C_i, \tau_i]$. We propose to model the event count to assess the effect of treatment on the rate of occurrence of events for the primary analysis. We further assume the primary analysis model parameter vector ϕ , which contains the treatment effect parameter, may be estimated in the complete data analysis via the estimating equations

$$\frac{1}{n} \sum_{i=1}^n \psi(Y_i, \phi) = 0. \quad (2)$$

2.2 | Primary analysis model

Popular count models for the primary analysis of recurrent events endpoints include the Poisson and negative binomial log-linear generalized linear models (GLMs). The negative binomial regression model can be more realistic for assessing the effect of treatment on the expected event rate. Unlike the Poisson regression model, but similar to the gamma-frailty model for the event intensity, the negative binomial regression model allows event rates between subjects to differ while assuming a constant within-subject rate of events. For the primary analysis, we thus assume the popular log-linear negative binomial regression model for the event count with log-offset of follow-up time:

$$\log\{E(Y_i|A_i, Z_i, C_i)\} = \alpha + A_i\delta + Z_i^T\beta + \log(C_i). \quad (3)$$

Here, $E(Y_i|A_i, Z_i, C_i)$ is the expected event count conditional on possible baseline characteristics, treatment assignment, and actual follow-up time. This model implicitly assumes data are MAR. The estimand of interest is δ , which entails the

treatment effect. The estimating equations (2) may be constructed for $\phi^T = (\alpha, \delta, \beta, \gamma)$, with a natural choice for $\psi(Y_i, \phi)$ being the score functions under model (3). In this model, γ is related to the dispersion parameter of the negative binomial regression.

2.3 | Imputation model

We now introduce the model for imputing missing events of subjects who drop out before the end of planned follow-up. For expository purposes, we assume the proportional intensity model with gamma frailty (1) and specify the parametric baseline intensity function

$$\Lambda_0(t, \lambda) = \lambda t. \quad (4)$$

In general, $\Lambda_0(t, \lambda)$ can be approximated by B-spline functions²⁰ to avoid specifying a precise parametric form for this function or to allow $\Lambda_0(t, \lambda)$ to be nonparametric. However, a benefit of the above linear specification considering the primary analysis model is that it corresponds to a count model for the expected number of events in follow-up time, t , given

$$\log\{E(Y_i|X_i, b_i, t)\} = \log\{\Lambda_{X_i}(t, b_i, \theta)\} = \alpha + A_i\delta + Z_i^T\beta + \log(t) + \log(b_i), \quad (5)$$

where $\theta^T = (\lambda, \delta, \beta, \gamma)$ and $\alpha = \log(\lambda)$. The gamma frailty leads to a marginal negative binomial distribution for the event count.¹⁶ While the intensity model (1) could be used to impute post-dropout event times,²¹ we instead impute the post-dropout event counts directly of interest for the primary analysis. The distribution of the event count post-dropout given the observed event count after integrating over b_i is

$$Y_{i,mis}|Y_{i,obs} \sim NB\left(k = \frac{1}{\gamma} + Y_{i,obs}, p = \frac{1 + \gamma\Lambda_0(C_i)e^{X_i^T\zeta}}{1 + \gamma\{\Lambda_0(\tau_i) - \Lambda_0(C_i)\}e^{\tilde{X}_i^T\zeta} + \gamma\Lambda_0(C_i)e^{X_i^T\zeta}}\right), \quad (6)$$

where $Y_{i,obs}$ is the observed number of events experienced through censoring time C_i . Here, k is the number of successes and p denotes the success probability. Control-based assumptions can be applied to this conditional distribution for the missing event count through the covariate vector, where X_i represents relevant covariates pre-dropout and \tilde{X}_i represents relevant covariates post-dropout. Covariates X_i and \tilde{X}_i will be further described in Section 4.

3 | MOTIVATING DATASETS

We will later demonstrate and evaluate our method utilizing both data from a real-world type I diabetes trial and a simulated dataset.

The Juvenile Diabetes Research Foundation conducted a “Randomized Clinical Trial to Assess the Efficacy of Real-Time Continuous Glucose Monitoring in the Management of Type I Diabetes” Jan 2007 to Jan 2009.²² The source of the data is the JDRF Artificial Pancreas Project sponsored by Jaeb Center for Health Research, but the analyses, content and conclusions presented herein are solely the responsibility of the authors and have not been reviewed or approved by Jaeb Center for Health Research. Though not the original primary endpoint, one safety outcome of interest was the rate of severe hypoglycemic events. For illustration, we will consider primary and sensitivity analyses of this endpoint utilizing models and imputation assumptions discussed herein.

In this trial, 232 subjects were randomized to real-time continuous glucose monitoring (CGM) with self-monitoring blood glucose and 211 were randomized to self-monitoring blood glucose alone (control) for 26 weeks. Subjects were followed in an extension study for an additional 26 weeks, during which all patients were assigned CGM. For the purpose of illustrating methods introduced in this article, we assume planned follow-up was $\tau_0 = 182$ days (26 weeks) and $\tau_1 = 364$ days (52 weeks) for control and CGM arms, respectively. There were a total of 73 observed severe hypoglycemic events across both arms within the follow-up times we consider. All subjects on the control arm were followed for the full 182 days of follow-up. The rate of dropout was 51.7% for CGM patients during the 364 days of follow-up, yielding an overall dropout rate of 27.1%.

We assume a log-linear negative binomial model with log-offset of follow-up time for primary analysis, but include only the treatment assignment as the baseline covariate, resulting in the model

$$\log\{\text{E}(Y_i|A_i, C_i)\} = \alpha + A_i\delta + \log(C_i), \quad (7)$$

where Y_i , A_i , and C_i represent the hypoglycemic event count during follow-up, the treatment assignment as previously defined, and the time of follow-up for subject i , respectively. Baseline covariates available for inclusion in the imputation model (1) include age, height, weight, sex, and duration of diabetes prior to enrollment, though we will discuss the selection of baseline covariates in greater detail in Section 8.

Now, in this trial, all subjects randomized to control were followed for the entire assumed duration of planned follow-up, which was shorter than that planned for the active treatment arm. We also simulate a motivating dataset in which we assume a randomized clinical trial design with equal allocation to active treatment and control, equal length of planned follow-up for all subjects, and similar dropout rates among both active treatment and control arms. We assume the endpoint of interest is again the event count, and incorporate the same baseline covariates in both imputation (1) and primary analysis (3) models for this dataset. The simulation and analysis of this dataset will be discussed in greater detail in Section 7. We will perform control-based sensitivity analyses on both the simulated and type I diabetes trial datasets utilizing DI for recurrent events.

4 | CONTROL-BASED ASSUMPTIONS

Control-based imputation was first proposed for missing data sensitivity analyses in clinical trials by Carpenter et al¹⁴ for continuous longitudinal data. Two approaches to control-based imputation that naturally extend to recurrent events are copy reference (CR) and jump to reference (J2R) imputation.²³ In CR imputation, subjects lost to follow-up on the active treatment arm are assumed to follow the observed distribution of the control arm subjects pre- and post-dropout. J2R imputation, on the other hand, is predicated on the assumption that subjects on the active treatment arm who are lost to follow-up “jump” to the observed control arm distribution only post-dropout. Under CR and J2R assumptions, the control arm is imputed assuming events post-randomization are MAR. Imputation under both assumptions is performed conditional on a subject’s baseline profile and history of past events.

Though it may seem counter-intuitive, J2R has been observed to produce more conservative treatment effect estimates than CR imputation for recurrent events.⁵ J2R is often conservative because the underlying assumption implies subjects will essentially lose any beneficial effect of their time on active treatment.⁴ Under J2R imputation, if a subject has a higher than average event rate prior to dropout, then their imputed event rate post-dropout will be higher than typically observed on the control arm because we condition on past history.²³ Under CR imputation, where subjects lost to follow-up are assumed to follow the observed control arm distribution of events for the full length of planned follow-up, a subject’s event rate prior to dropout is indicative of their personal event propensity. So, if a subject has a lower event rate pre-dropout than would be expected had they actually been randomized to the control arm, this prior event rate will feed into their post-dropout CR imputation. CR imputation can imitate when subjects lost to follow-up are essentially nonresponders.⁴ Though J2R can be more conservative, it is a reasonable strategy if the reference arm is the standard of care likely to be given to a subject after dropout,²⁴ such as in the JDRF type I diabetes trial.

We consider CR and J2R control-based imputation of missing events post-dropout. We also employ randomized-arm MAR (MAR) imputation,⁴ in which the post-dropout distribution of events for a subject is assumed to be the observed distribution for subjects on the arm to which they were actually randomized. MAR imputation can mimic the typical analysis strategy for the primary analysis. Control-based and MAR assumptions can be incorporated in the imputation model through model covariates used pre- and post-dropout, and events may then be imputed via the posterior distribution of events post-dropout (6). We specify relevant pre-and post-dropout covariates for subjects lost to follow-up on the active treatment arm:

- CR: $X_i^T = (A_i = 0, Z_i)$ pre-dropout and $\tilde{X}_i^T = (\tilde{A}_i = 0, Z_i)$ post-dropout;
- J2R: $X_i^T = (A_i, Z_i)$ pre-dropout and $\tilde{X}_i^T = (\tilde{A}_i = 0, Z_i)$ post-dropout; and
- MAR: $X_i^T = (A_i, Z_i)$ pre-dropout and $\tilde{X}_i^T = (\tilde{A}_i = A_i, Z_i)$ post-dropout.

5 | EXISTING METHODS: MULTIPLE IMPUTATION

A popular method for applying control-based assumptions in sensitivity analyses with recurrent events is Rubin's multiple imputation (MI).⁹ Let f be the imputation distribution dependent on imputation parameter θ . In the randomized clinical trials setting where imputation and analysis models are often proposed by the same party, it is common for θ and ϕ to contain overlap. Given the imputation and analysis models described in Section 2, the imputation parameter is $\theta^T = (\lambda, \delta, \beta, \gamma)$ and the analysis model parameter is $\phi^T = (\alpha, \delta, \beta, \gamma)$. The general steps for MI of the event count are as follows.

1. For $m \in \{1, \dots, M\}$, obtain an estimate, $\hat{\theta}^{(m)}$, of the imputation parameter, θ , given the observed data and imputation model, f . Impute missing post-dropout event count $Y_{i,mis}^{*(m)}$ under the imputation model, f , conditional on $Y_{i,obs}$ at this realization of $\hat{\theta}^{(m)}$ to create M complete datasets.
2. Perform the primary analysis under the proposed analysis model on each of the M completed datasets to obtain estimates of the analysis model parameter, $\hat{\phi}^{(m)}$.

Rubin's combining rules⁹ are used to arrive at the MI estimator for the analysis parameter, $\hat{\phi}_{MI}$, and its variance:

$$\hat{\phi}_{MI} = \frac{1}{M} \sum_{m=1}^M \hat{\phi}^{(m)}, \quad V(\hat{\phi}_{MI}) = \bar{W}_M + (1 + M^{-1})B_M. \quad (8)$$

Here, $\bar{W}_M = M^{-1} \sum_{m=1}^M W_m$ is the average within-imputation variance, W_m is the estimated variance of $\hat{\phi}^{(m)}$, and the between-imputation variance, B_M , is the sample variance of the $\hat{\phi}^{(m)}$'s. The complete event count for partially observed subject i is constructed: $Y_i^{*(m)} = Y_{i,obs} + Y_{i,mis}^{*(m)}$.

The typical approach in the literature for the selection of $\hat{\theta}^{(m)}$ in missing recurrent events sensitivity analyses with control-based MI is to sample the imputation model parameter via either Bayesian posterior draw methods^{16,23} or from the asymptotic distribution of the maximum likelihood estimator (MLE).^{5,21} A more computationally efficient option, however, is to estimate θ as the pseudo-MLE from observed data (MLMI), since the MLE is the same in every imputation iteration.¹⁷ MLMI is appealing due to its relative computational efficiency, but Rubin's variance estimator can be biased for MLMI²⁵ because imputing missing outcomes from the conditional distribution given the observed data evaluated at the MLE is "improper."⁹

Regardless of the approach to MI, Rubin's variance estimator has been shown to be biased for the true sampling variance of the MI estimator in applications of control-based imputation of recurrent events.²⁶ While MI can provide valid inference for data MAR,²⁷ this is not always true for data MNAR or in cases of uncongeniality or model misspecification.²⁸ Congeniality generally means the imputation and analysis model classes are compatible (see Xie and Meng²⁹ and Meng³⁰ for a formal definition). This congeniality condition is problematic in our control-based imputation setting, as control-based assumptions made in sensitivity analyses retaining the primary analysis model impose uncongeniality.³¹ Control-based imputation is predicated on the assumption that data are MNAR. Even if imputation and analysis models are of similar forms, performing imputation assuming events are MNAR results in incompatibility between the imputation and primary analysis models. Despite Rubin's variance estimator being known to be biased in this setting, conventional use of Rubin's combining rules following control-based MI of recurrent events remains popular in the literature.^{21,23}

Bootstrapping methods offer flexibility in improving variance estimation of the MI point estimator. The nonparametric bootstrap has been successfully utilized to estimate the variance of treatment effect estimators with control-based MI of recurrent events,⁵ and has been shown to yield more accurate estimates of standard error and proper coverage for confidence intervals under control-based MI of recurrent events in comparison to Rubin's variance estimator.¹⁶ Though it offers improvements over Rubin's variance estimator, the nonparametric bootstrap can be very computationally intensive regardless of the method used to select $\hat{\theta}^{(m)}$. This can be a major disadvantage in practice. Given the deficiencies of MI with Rubin's variance estimator in the control-based setting and practical limitations of other popular methods of variance estimation for the MI point estimator, we propose distributional imputation (DI) with a parallel wild bootstrap variance estimation procedure for missing recurrent events sensitivity analyses.

6 | DISTRIBUTIONAL IMPUTATION FOR RECURRENT EVENTS

6.1 | Distributional imputation procedure

We propose DI with wild bootstrap variance estimating procedure for estimating and assessing the uncertainty of the treatment effect in control-based missing data sensitivity analyses of recurrent events endpoints. Liu et al³² also recently suggested DI for performing control-based sensitivity analyses in the continuous longitudinal data setting. Liu et al demonstrated via simulations that DI produces comparable point estimates to MI for a variety of estimands and outperforms Rubin's combining rules regarding coverage probabilities and relative bias of estimated variance in the continuous longitudinal data setting.³² However, the DI method proposed by Liu et al only handles continuous and binary data. We propose a DI method for recurrent events sensitivity analyses.

Recall we assume the target analysis parameter, ϕ , may be estimated by solving $n^{-1} \sum_{i=1}^n \psi(Y_i, \phi) = 0$. In the presence of missing data and under MAR, a consistent estimator for ϕ may be obtained by solving

$$n^{-1} \sum_{i=1}^n E\{\psi(Y_i, \phi) | Y_{i,obs}, \hat{\theta}\} = 0. \quad (9)$$

This expectation is with respect to imputation density $f(Y_{i,mis} | Y_{i,obs}, \theta)$, which in our case is the conditional negative binomial distribution (6). Like in MLMI, the imputation parameter θ may be estimated by the pseudo-MLE, $\hat{\theta}$, via the mean score equations, $n^{-1} \sum_{i=1}^n E\{s(Y_i, \theta) | Y_{i,obs}\} = 0$, given the imputation model.

Rooted in the idea of Monte Carlo (MC) integration,³³ the conditional expectation estimating equations for ϕ (9) may be approximated by $M^{-1} \sum_{m=1}^M \psi(Y_i^{*(m)}, \phi)$, where $Y_i^{*(m)}$ is the complete data post-imputation. Thus, ϕ may be estimated by solving:

$$\frac{1}{nM} \sum_{i=1}^n \sum_{m=1}^M \psi(Y_i^{*(m)}, \phi) = 0. \quad (10)$$

The procedure for DI is then as follows.

1. Calculate the imputation model estimator, $\hat{\theta}$, from the observed data via the mean score equations under the imputation model.
2. For $m \in \{1, \dots, M\}$, impute missing event counts $Y_{i,mis}^{*(m)} \sim f(Y_{i,mis} | Y_{i,obs}, \hat{\theta})$.
3. Obtain the DI estimator for the analysis model parameter, $\hat{\phi}_{DI}$, by solving (10).

A fundamental difference between DI and MI is that $\hat{\phi}_{DI}$ is estimated using all of the completed data, rather than by averaging over the analysis performed in M single complete datasets. We note that estimating $\hat{\phi}_{DI}$ from the pooled M completed datasets is similar to what is done under parametric fractional imputation (FI).^{34,35} In contrast to FI, DI does not require importance sampling or a proposal distribution, but takes advantage of the estimated conditional distribution given observed data under the control-based sensitivity assumptions for direct imputation of missing data. This makes DI more straightforward to implement with control-based assumptions. Pooling over the completed datasets can make estimation of the point estimator more computationally efficient under DI than MI, but can also require more storage than MI. To reduce storage costs of DI, a weighted regression technique can be utilized for the estimating equations (10) so only one copy of the event count must be retained for subjects followed to completion.

6.2 | Variance estimation

The variance of the point estimator obtained under DI may be estimated via an accompanying wild bootstrap procedure,³⁶ which parallels the DI procedure and draws on the concept of importance sampling³⁷ to account for the variability contributed from estimating both the imputation and analysis parameters. This wild bootstrap, like the non-parametric bootstrap, is flexible for easy application to the recurrent events setting. The wild bootstrap procedure is as follows.

1. Randomly generate i.i.d. wild bootstrap weights, $u_i^{(b)}$, with mean and variance 1 such that $u_i^{(b)} \geq 0$ for each subject i . Here b denotes the current bootstrap replicate. Calculate the imputation parameter estimate replicate $\hat{\theta}^{(b)}$ by solving the weighted mean score equations:

$$\frac{1}{n} \sum_{i=1}^n u_i^{(b)} E\{s(Y_i, \theta) | Y_{i,obs}\} = 0. \quad (11)$$

2. Update, importance weights $w_i^{(m)}(\hat{\theta}^{(b)})$, subject to $\sum_{m=1}^M w_i^{(m)}(\hat{\theta}^{(b)}) = 1$ for all i , where

$$w_i^{(m)}(\hat{\theta}^{(b)}) \propto \frac{f(Y_i^{*(m)} | Y_{i,obs}, \hat{\theta}^{(b)})}{f(Y_i^{*(m)} | Y_{i,obs}, \hat{\theta})}. \quad (12)$$

3. Obtain the DI estimate replicate, $\hat{\phi}_{DI}^{(b)}$, by solving the weighted estimating equations:

$$\frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M u_i^{(b)} w_i^{(m)}(\hat{\theta}^{(b)}) \psi(Y_i^{*(m)}, \phi) = 0. \quad (13)$$

We repeat steps 1-3 for $b \in \{1, \dots, B\}$. Then the variance of $\hat{\phi}_{DI}$ may be estimated by

$$\hat{V}_{WB}(\hat{\phi}_{DI}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\phi}_{DI}^{(b)} - \hat{\phi}_{DI})^{\otimes 2}. \quad (14)$$

By constructing the importance weights subject to the conditions given in Step 2, (13) approximates the bootstrap replication of the estimating equations (9) without re-imputation of the missing values. Solving these weighted estimating equations (13) for $\hat{\phi}^{(b)}$ involves the expectation conditional on both the current bootstrap sample and the estimated $\hat{\theta}^{(b)}$. Thus, the importance weights are constructed to account for the variability introduced by estimating the imputation parameter in the current bootstrap sample. Furthermore, control-based assumptions can be easily implemented through updating these importance weights (12) using the conditional distribution of events given the observed event count.

Now, there are many options for the distribution of the wild bootstrap weights, $u_i^{(b)}$, including exponential or Poisson distributions with rate parameter 1. The procedure is not sensitive to the choice of wild bootstrap weight distribution, given the conditions for $u_i^{(b)}$ specified in Step 1 above.³² In comparison to the nonparametric bootstrap, this wild bootstrap procedure offers a gain in computational efficiency by avoiding repeating the full imputation and analysis processes in each bootstrap iteration.

6.3 | Asymptotic results

The DI estimator and corresponding variance estimator based on the wild bootstrap procedure exhibit favorable asymptotic properties, which we discuss in this section. We present theorems for the consistency and asymptotic normality of the DI point estimator, $\hat{\phi}_{DI}$, and for the consistency of the variance estimator, $\hat{V}_{WB}(\hat{\phi}_{DI})$, based on the wild bootstrap procedure. Proofs of all presented theorems are given in the Supporting Information.

Theorem 1 (consistency of $\hat{\phi}_{DI}$). *Under regularity conditions given in Section A.1 of the Supporting Information, $\hat{\phi}_{DI}$ converges in probability to ϕ_0 as sample size $n \rightarrow \infty$ and imputation size $M \rightarrow \infty$, where ϕ_0 is the true value of the target analysis parameter.*

Theorem 2 (asymptotic normality of $\hat{\phi}_{DI}$). *Under regularity conditions given in Section A.2 in the Supporting Information, and as sample size $n \rightarrow \infty$ and imputation size $M \rightarrow \infty$,*

$$\sqrt{n}(\hat{\phi}_{DI} - \phi_0) \text{ converges in distribution to } N(0, V_{DI}),$$

where $V_{DI} = E(K_1 K_1^T)$ and

$$K_i = C_2 E\{\psi(Y_i, \phi_0) | Y_{i,obs}, \theta_0\} + C_2 C_1 C_3 E\{s(Y_i, \theta_0) | Y_{i,obs}\};$$

$$C_1 = E[\{s(Y_i, \theta_0) - s_{obs}(Y_i, \theta_0)\}\psi(Y_i, \phi_0)]; \quad C_2 = -\left[E\left\{\frac{\partial}{\partial \theta}\psi(Y_i, \phi_0)\right\}\right]^{-1}; \quad C_3 = I_{\text{obs}}^{-1}(\theta_0).$$

$I_{\text{obs}}(\theta_0)$ is the observed Fisher information of the true value of the imputation parameter, θ_0 , and $s_{obs}(Y_i, \theta_0)$ is the observed score equation with respect to the imputation model evaluated at θ_0 .

Theorem 3 (consistency of $\hat{V}_{WB}(\hat{\phi}_{DI})$). *Let wild bootstrap weights $u_i^{(b)} \geq 0$ for all $i = 1, \dots, n$ be i.i.d. with mean and variance 1. Then under regularity conditions given in Section A.3 of the Supporting Information, $\hat{V}_{WB}(\hat{\phi}_{DI})$ converges in probability to \hat{V}_{DI} as sample size $n \rightarrow \infty$, imputation size $M \rightarrow \infty$, and bootstrap size $B \rightarrow \infty$.*

Concerning the choice of imputation size M and bootstrap size B , larger values of both offer improved performance if computationally feasible. However, results from our simulation studies show that performance of the DI point and variance estimators for the treatment effect is not very susceptible to the choice of M .

7 | SIMULATION STUDY

We design a simulation study similar to that given by Gao et al¹⁶ to evaluate the finite-sample performance of DI with the wild bootstrap for the treatment effect (δ_{DI}) under CR, J2R, and MAR imputation. We compare the point and variance estimators given by DI to those for the treatment effect produced by MI (δ_{MI}) with either Rubin's combining rules or a nonparametric bootstrap. We consider a randomized clinical trial design with total sample size $n \in \{200, 2000\}$ with equal allocation to active treatment and control arms. The outcome of interest is the event count during specified follow-up period $\tau_i = 5$ for all $i \in \{1, \dots, n\}$. We utilize the log-linear negative binomial regression model (3) for the planned analysis to estimate treatment effect, δ , where the full analysis parameter is $\phi^T = (\alpha, \delta, \beta, \gamma)$.

We consider performance across 1000 simulated datasets, generating event counts assuming a gamma frailty model with cumulative intensity function (1), linear baseline intensity function (4), and initial parameter $\theta_0^T = (\lambda = 0.5, \delta_0 = -0.8, \beta = 0.5, \gamma = 1)$. We specify initial δ_0 , as the true treatment effect δ is expected to change under varying control-based assumptions and rates of missingness. The true values of δ are estimated by performing MI and DI with $n = 10,000$ and $M = 100$, and are shown in Table 1. We include one baseline covariate, $Z_i \stackrel{i.i.d.}{\sim} \text{Unif}(0, 1)$, in the data generation, imputation, and analysis models.

We assume monotone missingness due to loss to follow-up. We generate a noninformative censoring time, $C_i \leq \tau_i$, representing the time of dropout or completion of planned follow-up according to $C_i = B_i \tau_i + (1 - B_i) \text{Unif}(0, \tau_i)$, where $B_i \stackrel{i.i.d.}{\sim} \text{Bernoulli}(p_0)$. We vary $p_0 \in \{0.8, 0.5, 0.3\}$, corresponding to expected dropout rates (DORs) of $E(\text{DOR}) \in \{20\%, 50\%, 70\%\}$. The rate of missing events corresponds to approximately $E(\text{DOR})/2$ here, shown empirically in Table 2.

We assume the intensity model with gamma frailty and linear baseline intensity function to avoid model misspecification in the imputation process. Initial estimation of imputation parameter $\theta^T = (\lambda, \delta, \beta, \gamma)$ is performed utilizing all of the observed data under J2R and MAR imputation and utilizing only the observed control arm data under CR imputation. Missing event counts post-dropout are imputed from the conditional negative binomial density (6). We consider imputation size $M \in \{5, 50\}$. For the bootstrapping methods, we utilize bootstrap size $B = 200$. For the wild bootstrap variance estimating procedure, we assume wild bootstrap weights $u_i \stackrel{i.i.d.}{\sim} \text{Exp}(1)$ for $i \in \{1, \dots, n\}$.

TABLE 1 True values of δ_{MI} and δ_{DI} for each imputation assumption and expected dropout rate ($E(\text{DOR})$).

E(DOR)	CR			J2R			MAR		
	20%	50%	70%	20%	50%	70%	20%	50%	70%
δ_{MI}	-0.735	-0.644	-0.588	-0.684	-0.533	-0.443	-0.800	-0.800	-0.800
δ_{DI}	-0.734	-0.644	-0.588	-0.684	-0.533	-0.443	-0.800	-0.800	-0.800

TABLE 2 Observed dropout rates (DORs) and missing event rates (MRs) by E(DOR).

Rate	<i>n</i> = 200			<i>n</i> = 2000		
	20%	50%	70%	20%	50%	70%
DOR	0.200	0.498	0.699	0.200	0.500	0.700
MR	0.101	0.249	0.350	0.100	0.251	0.351

Note: *n* = total sample size.

We assess performance of DI with the wild bootstrap for variance estimation in comparison to MI with either Rubin's combining rules or a nonparametric bootstrap under CR, J2R, and MAR imputation in terms of treatment effect point estimator bias, relative bias of estimated standard errors (SEs), average confidence interval lengths and average computational time, presented in Table 3. Confidence intervals are standard Wald intervals. MI and DI produce similar estimates of δ in simulations, with both estimators tending to be negatively biased and $\hat{\delta}_{DI}$ typically being less than or as biased as $\hat{\delta}_{MI}$. The bias of either estimator improves with increases in sample size, *n*. The magnitude of these biases tends to decrease as dropout rate increases under CR and J2R imputation, which is likely due to the decrease in magnitude of treatment effect estimates as dropout rate increases.

The true standard errors of $\hat{\delta}_{MI}$ and $\hat{\delta}_{DI}$ are similar. DI with the wild bootstrap offers great improvements over MI with Rubin's combining rules for accurate estimation of the true sampling variability of the point estimator under CR and J2R imputation. For these control-based assumptions, the relative biases of the estimated standard error under DI with the wild bootstrap are more similar to those produced by MI with the nonparametric bootstrap, with the latter typically yielding slightly less biased standard error estimates for the treatment effect. Under J2R and CR assumptions, relative bias remains similar despite increasing sample size, and performance of estimated standard error is more accurate under lower dropout rates for either estimator. Relative bias of the standard error estimates for $\hat{\delta}_{MI}$ with Rubin's combining rules under CR and J2R imputation tends to increase as imputation size *M* increases, whereas those for $\hat{\delta}_{DI}$ or MI with the nonparametric bootstrap are affected to a lesser degree under changes to *M*. In general, relative biases for $\hat{\delta}_{DI}$ and $\hat{\delta}_{MI}$ with the nonparametric bootstrap are much less varied than that for $\hat{\delta}_{MI}$ with Rubin's combining rules. This suggests DI with the wild bootstrap, similar to the nonparametric bootstrap when used with MI, is less susceptible to changes in sample size, imputation size, or dropout rates than MI with Rubin's combining rules in more consistently estimating standard error of the treatment effect estimator. While DI with the wild bootstrap more accurately estimates the standard error of δ than MI with Rubin's combining rules under the control-based assumptions considered, this is not necessarily the case under MAR imputation. MI typically, if not uniformly, appears to produce more accurate standard error estimates with either variance estimator under MAR imputation than DI with the wild bootstrap procedure. DI does experience improvements in standard error estimation under MAR imputation with increases in *M* and sample size.

One trend we observe is that the true standard error decreases for both point estimators as dropout rate increases in the control-based imputation scenarios. While this may seem counter-intuitive, it is likely due in part to the decreasing magnitude of δ for increasing dropout rates under control-based imputation assumptions. Additionally, Xie and Meng²⁹ found that in some cases the efficiency of MI estimators increases as the amount of missingness increases. For example, in this setting, the score equations used in estimating $\hat{\delta}_{MI}$ and $\hat{\delta}_{DI}$ are not the true observed score equations, as obtaining the latter would require taking the expectation of the former with respect to the observed data. Doing so is not always practically feasible, and could create an added barrier to use of these methods. Because we do not take this expectation, the obtained estimator under CR and J2R imputation is no longer the true MLE, and may not be the most efficient estimator for δ . Despite this, the wild bootstrap for $\hat{\delta}_{DI}$, like the nonparametric bootstrap for $\hat{\delta}_{MI}$, is able to reflect this true trend much better than Rubin's combining rules for variance estimation of $\hat{\delta}_{MI}$.

As we expect, MI with Rubin's combining rules is very computationally efficient. Though DI with the wild bootstrap is computationally more intensive than MI with Rubin's combining rules, DI is still computationally more efficient than MI with the nonparametric bootstrap. Average computational time increases for all three estimation methods as sample size and *M* increase. Average computational times for MI with either variance estimating procedure appear affected to a greater degree by increases in *M* as compared to sample size, while the opposite appears to be the case for DI. Average computational time for DI with the wild bootstrap increases with increasing dropout rates, whereas MI with either variance estimator is less affected by changes to dropout rates. Thus, DI with the wild bootstrap offers increased improvements to computation time over MI with the nonparametric bootstrap under low to moderate dropout rates.

TABLE 3 Performance of DI with wild bootstrap, MI with Rubin's combining rules, and MI with nonparametric bootstrap under CR, J2R, and MAR imputation with $n \in \{200, 2000\}$, $M \in \{5, 50\}$, and expected dropout rate $E(DOR) \in \{20\%, 50\%, 70\%\}$ for treatment effect δ .

$\hat{\delta}_{MI}$		Rubin's rules										Nonparametric bootstrap										Wild Bootstrap									
		E	M	(DOR)	True	\widehat{SE}	CI length	Time (s)	Bias ^a	SE	\widehat{SE}	RB ^b	CI length	Time (s)	Bias ^a	SE	\widehat{SE}	RB ^b	CI length	Time (s)											
<i>Copy reference imputation</i>																															
200	5	0.2	-8	0.164	0.182	1.115	0.71	0.1	-8	0.164	0.164	1.001	0.64	16.2	-7	0.163	0.160	0.979	0.63	6.1											
200	5	0.5	-8	0.154	0.193	1.255	0.76	0.1	-8	0.154	0.151	0.982	0.59	16.0	-6	0.154	0.148	0.962	0.58	8.1											
200	5	0.7	-5	0.150	0.200	1.330	0.78	0.1	-5	0.150	0.145	0.961	0.57	16.0	-3	0.150	0.142	0.946	0.56	9.3											
200	50	0.2	-8	0.163	0.181	1.112	0.71	0.8	-8	0.163	0.162	0.992	0.63	141.4	-7	0.163	0.157	0.966	0.62	19.7											
200	50	0.5	-6	0.150	0.191	1.272	0.75	0.8	-6	0.150	0.147	0.981	0.58	138.8	-4	0.149	0.143	0.957	0.56	40.2											
200	50	0.7	-5	0.145	0.196	1.350	0.77	0.8	-5	0.145	0.139	0.955	0.54	137.6	-2	0.144	0.134	0.932	0.53	53.3											
2000	5	0.2	-4	0.051	0.058	1.134	0.23	0.4	-4	0.051	0.052	1.012	0.20	82.2	-4	0.051	0.052	1.016	0.20	34.2											
2000	5	0.5	-2	0.049	0.061	1.247	0.24	0.4	-2	0.049	0.048	0.971	0.19	82.9	-2	0.049	0.048	0.977	0.19	51.8											
2000	5	0.7	-2	0.047	0.063	1.337	0.25	0.4	-2	0.047	0.046	0.967	0.18	83.6	-2	0.047	0.046	0.976	0.18	63.6											
2000	50	0.2	-4	0.051	0.058	1.128	0.23	0.8	-4	0.051	0.051	1.004	0.20	830.2	-4	0.051	0.051	1.002	0.20	157.8											
2000	50	0.5	-2	0.048	0.061	1.266	0.24	0.8	-2	0.048	0.046	0.971	0.18	823.0	-2	0.048	0.046	0.969	0.18	360.1											
2000	50	0.7	-2	0.045	0.062	1.372	0.24	0.8	-2	0.045	0.044	0.970	0.17	822.5	-2	0.045	0.044	0.968	0.17	491.9											
<i>Jump to reference imputation</i>																															
200	5	0.2	-7	0.149	0.185	1.241	0.72	0.1	-8	0.150	0.150	1.001	0.59	17.5	-6	0.149	0.150	1.007	0.59	7.5											
200	5	0.5	-6	0.124	0.196	1.577	0.77	0.1	-8	0.123	0.122	0.996	0.48	17.3	-4	0.124	0.126	1.016	0.49	9.5											
200	5	0.7	-5	0.108	0.199	1.832	0.78	0.1	-5	0.109	0.107	0.977	0.42	17.3	-3	0.108	0.113	1.044	0.44	10.7											
200	50	0.2	-8	0.148	0.183	1.242	0.72	0.8	-9	0.148	0.148	0.996	0.58	141.7	-6	0.147	0.144	0.980	0.57	20.8											
200	50	0.5	-7	0.118	0.193	1.633	0.76	0.8	-7	0.118	0.116	0.988	0.46	137.3	-4	0.117	0.114	0.970	0.45	40.6											
200	50	0.7	-6	0.101	0.196	1.940	0.77	0.8	-5	0.101	0.098	0.973	0.38	135.2	-2	0.100	0.097	0.963	0.38	53.0											
2000	5	0.2	-4	0.047	0.059	1.252	0.23	0.5	-5	0.046	0.047	1.016	0.19	91.0	-4	0.047	0.048	1.027	0.19	43.2											
2000	5	0.5	-1	0.039	0.062	1.584	0.24	0.5	-2	0.039	0.038	0.971	0.15	93.1	-1	0.039	0.040	1.031	0.16	62.2											
2000	5	0.7	-3	0.034	0.063	1.861	0.25	0.5	-2	0.034	0.033	0.967	0.13	94.3	-2	0.034	0.036	1.059	0.14	74.6											
2000	50	0.2	-4	0.046	0.058	1.264	0.23	3.9	-4	0.046	0.046	1.012	0.18	837.8	-4	0.046	0.047	1.011	0.18	166.3											
2000	50	0.5	-2	0.037	0.061	1.632	0.24	3.9	-2	0.037	0.036	0.970	0.14	829.0	-1	0.037	0.036	0.976	0.14	367.9											
2000	50	0.7	-2	0.031	0.062	1.973	0.24	3.9	-2	0.031	0.030	0.969	0.12	825.3	-2	0.031	0.031	0.977	0.12	498.6											

(Continues)

TABLE 3 (Continued)

n	M	E(DOR)	Rubin's rules				Nonparametric bootstrap				Wild Bootstrap			
			True	SE	CI length	Time (s)	Bias ^a	SE	CI length	Time (s)	Bias ^a	SE	CI length	Time (s)
<i>Missing at random imputation</i>														
200	5	0.2	-12	0.182	0.179	0.983	0.70	0.1	-12	0.181	0.181	0.996	0.71	17.8
200	5	0.5	-14	0.195	0.189	0.967	0.74	0.1	-14	0.194	0.193	0.993	0.76	18.2
200	5	0.7	-16	0.210	0.195	0.929	0.76	0.1	-16	0.212	0.203	0.957	0.79	18.6
200	50	0.2	-12	0.181	0.178	0.985	0.70	0.8	-12	0.181	0.180	0.991	0.70	145.4
200	50	0.5	-14	0.193	0.187	0.969	0.73	0.8	-13	0.192	0.191	0.992	0.75	146.8
200	50	0.7	-16	0.208	0.193	0.928	0.76	0.8	-16	0.208	0.199	0.961	0.78	148.0
2000	5	0.2	-5	0.057	0.057	1.002	0.22	0.5	-5	0.057	0.057	1.002	0.22	91.9
2000	5	0.5	-2	0.062	0.060	0.964	0.23	0.5	-3	0.061	0.061	0.988	0.24	94.8
2000	5	0.7	-5	0.064	0.062	0.957	0.24	0.5	-4	0.064	0.063	0.994	0.25	97.0
2000	50	0.2	-5	0.057	0.057	1.000	0.22	3.9	-5	0.057	0.057	1.001	0.22	848.0
2000	50	0.5	-2	0.061	0.059	0.968	0.23	3.9	-2	0.061	0.060	0.981	0.23	849.9
2000	50	0.7	-4	0.063	0.061	0.958	0.24	3.9	-4	0.063	0.062	0.988	0.24	856.4

^aBias is presented $\times 10^3$.^bRB = Relative Bias (Ratio) of Estimated SE/True SE.

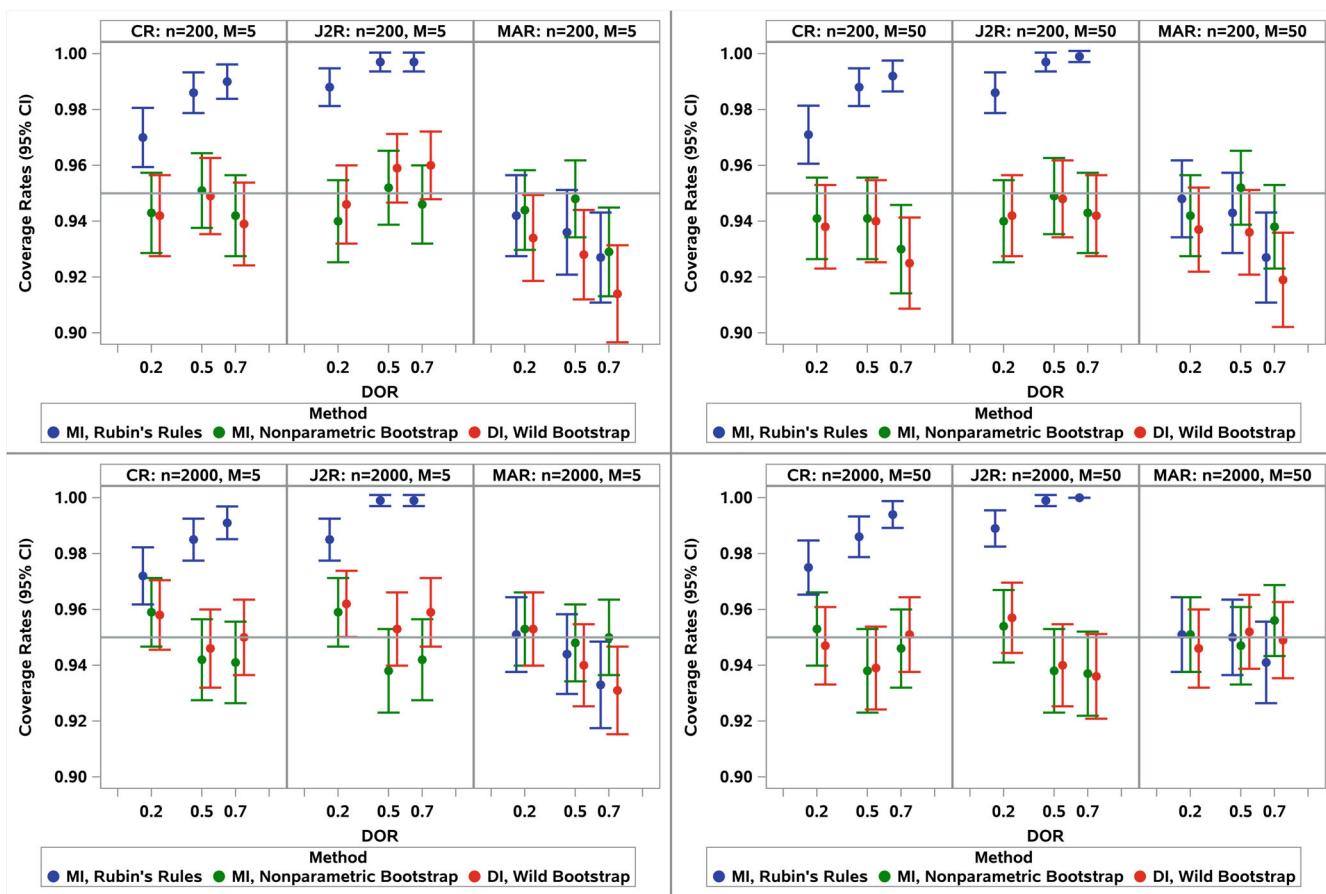


FIGURE 1 Estimated coverage rates for Wald confidence intervals of $\hat{\delta}_{MI}$ with Rubin's rules (blue), $\hat{\delta}_{MI}$ with nonparametric bootstrap (green), and $\hat{\delta}_{DI}$ (red) under CR, J2R, and MAR imputation with sample size $n \in \{200, 2000\}$ and imputation size $M \in \{5, 50\}$.

In addition to the confidence interval lengths given in Table 3, estimated coverage rates are presented with 95% confidence limits in Figure 1. DI with the wild bootstrap and MI with the nonparametric bootstrap tend to produce shorter confidence intervals than MI with Rubin's combining rules. DI with the wild bootstrap and MI with the nonparametric bootstrap also produce coverage rates typically closer to the nominal level than MI with Rubin's combining rules under CR and J2R imputation, though 95% confidence intervals for the estimated coverage of $\hat{\delta}_{DI}$ and $\hat{\delta}_{MI}$ when utilizing the nonparametric bootstrap fall completely below the nominal level under CR imputation for the smaller sample size and highest dropout rate. Under MAR imputation, the performance of MI and DI are more similar. However, MI with either variance estimation method often produces confidence intervals with coverage rates closer to the nominal level than DI under MAR imputation, particularly for smaller sample sizes.

We additionally assess type I error rates of these methods setting $\delta_0 = 0$, which yields true δ values very close to zero. DI with the wild bootstrap and MI with the nonparametric bootstrap similarly produce estimated type I error rates (displayed with 95% confidence limits in Figure 2) closer to the nominal level, which generally improve as sample size increases and dropout rate decreases, than MI with Rubin's combining rules under CR and J2R imputation. For MAR imputation, however, MI results in improved type I error rates over DI under smaller M and sample size. MI with the nonparametric bootstrap on average offers the best performance for type I error rates under MAR imputation.

We also evaluate the power of $\hat{\delta}_{DI}$ and $\hat{\delta}_{MI}$ when using Rubin's combining rules, displayed in Figure 3, as the magnitude of initial δ_0 increases. We exclude MI with the nonparametric bootstrap from the assessment of power due to the high computational time of the nonparametric bootstrap. Under CR and J2R assumptions, estimated power increases much more rapidly with increases to the magnitude of δ_0 for $\hat{\delta}_{DI}$ than for $\hat{\delta}_{MI}$ with Rubin's combining rules, with estimated power improving for both estimators with increases to sample size and M and decreases to dropout rate. Estimated power for $\hat{\delta}_{MI}$ with Rubin's combining rules and $\hat{\delta}_{DI}$ is more similar under MAR imputation, with the most noticeable improvements resulting from increases to sample size. As we expect, estimated power improves most slowly under J2R imputation with

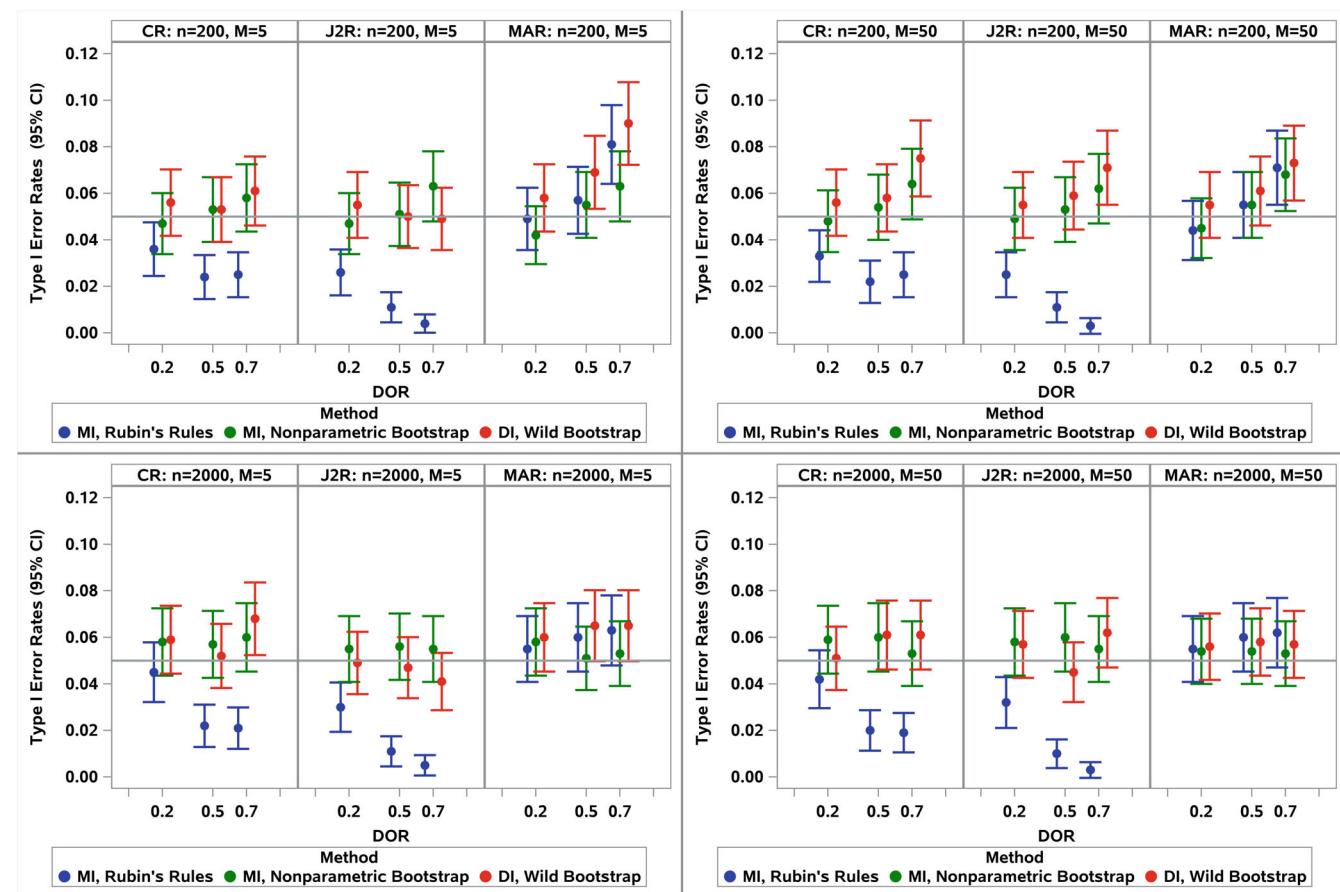


FIGURE 2 Estimated type I error rates for $\hat{\delta}_{MI}$ with Rubin's rules (blue), $\hat{\delta}_{MI}$ with nonparametric bootstrap (green), and $\hat{\delta}_{DI}$ (red) under CR, J2R, and MAR imputation with sample size $n \in \{200, 2000\}$ and imputation size $M \in \{5, 50\}$.

increasing magnitude of δ_0 , as we observe that J2R imputation produces the most conservative estimates of the treatment effect.

In these simulations, since we utilize the same pseudo-maximum likelihood-based methods for estimation of imputation parameter θ in our approach to MI as in our implementation of DI, we employ MLMI. These results empirically demonstrate that DI with the wild bootstrap more accurately estimates point estimator standard errors than MI with Rubin's combining rules when ML methods are used in estimating θ under control-based assumptions that introduce uncongeniality between imputation and analysis models. However, since MI with Rubin's combining rules in many cases outperforms DI under MAR imputation in these simulations, we cannot yet suggest that DI corrects the shortcomings of MLMI due only to its improper nature.

8 | DATA APPLICATION

We return to the motivating example, a “Randomized Clinical Trial to Assess the Efficacy of Real-Time Continuous Glucose Monitoring in the Management of Type I Diabetes.”²² Again, the source of the data is the JDRF Artificial Pancreas Project sponsored by Jaeb Center for Health Research, but the analyses, content and conclusions presented herein are solely the responsibility of the authors and have not been reviewed or approved by Jaeb Center for Health Research. Recall that we assume a log-linear negative binomial model with log-offset of follow-up time with treatment assignment as the only baseline covariate for analysis of the rate of hypoglycemic events:

$$\log\{E(Y_i|A_i, C_i)\} = \alpha + A_i\delta + \log(C_i), \quad (15)$$

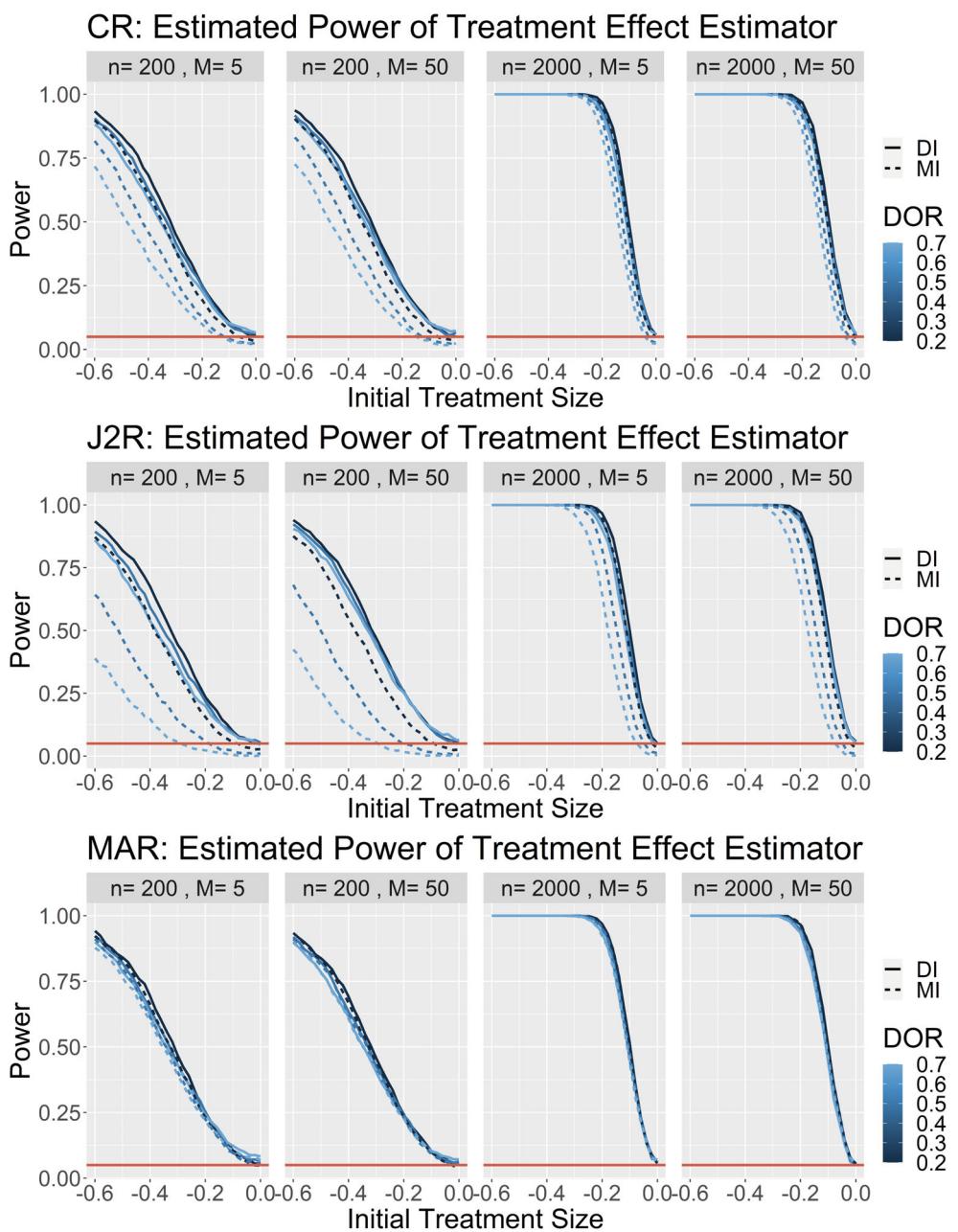


FIGURE 3 Estimated power for $\hat{\delta}_{MI}$ with Rubin's Rules (dashed) and $\hat{\delta}_{DI}$ (solid) with initial treatment effect size, $-0.6 \leq \delta_0 \leq 0$, under CR, J2R, and MAR imputation with sample size $n \in \{200, 2000\}$ and imputation size $M \in \{5, 50\}$.

where Y_i , A_i , and C_i represent the hypoglycemic event count during follow-up, the treatment assignment as previously defined, and the time of follow-up for subject i , respectively. With this model we can estimate the average treatment effect. For imputation, we assume a gamma frailty intensity model with linear baseline intensity, leading to the conditional negative binomial imputation density (6). It is advisable to select an imputation model at least as saturated as the primary analysis model,²⁹ but including too many auxiliary variables in the imputation model can result in poor model fit.²⁷ Considering this, we include in the imputation model additional baseline covariates, selected as those forming the corresponding negative binomial count model with the smallest AIC among those with at least three covariates other than A_i . We assume a treatment arm-specific imputation model, allowing baseline covariates to differ between CGM and control arms. Covariates considered for inclusion in the imputation model were age, height, weight, sex, and duration of diabetes prior to enrollment.

For fitting the imputation model, weight (W), height (H), and duration of diabetes (D) were selected for subjects on the CGM arm. For subjects on the control arm, weight (W), sex (S), and duration of diabetes (D) were included in fitting the imputation model. Thus, we fit the J2R and MAR imputation model with gamma frailty:

$$\Lambda(t; b_i, \lambda, \delta, \beta) = \lambda t b_i \exp\{\delta A_i + \beta_1 W_i + \beta_2 A_i H_i + \beta_3 (1 - A_i) S_i + \beta_4 D_i\}, \quad (16)$$

where b_i is an i.i.d. gamma distributed random variable with mean 1 and variance γ . We also fit the CR imputation model with gamma frailty using only the control data:

$$\Lambda(t; b_i, \lambda, \delta, \beta) = \lambda t b_i \exp(\delta A_i + \beta_1 W_i + \beta_2 S_i + \beta_3 D_i). \quad (17)$$

In addition to DI with the wild bootstrap and MLMI with Rubin's combining rules, we choose to demonstrate MI with the nonparametric bootstrap using the Bayesian imputation method given in Gao et al.¹⁶ This latter approach to imputation through Bayesian data augmentation methods and a nonparametric bootstrap was shown to lead to more accurate standard error estimates and proper coverage for confidence intervals under control-based imputation compared to those calculated with Rubin's variance estimator.¹⁶ Gao et al¹⁶ proposed a piece-wise exponential baseline intensity function for imputation of recurrent events. We specify $K = 1$ cutpoints for this piece-wise baseline intensity function, which results in the linear baseline intensity we specify previously. A sensitivity analysis for the number of K cutpoints used is given in the Supporting Information. We select noninformative priors for imputation model parameters, and set burn-in to 5000 and thinning to 100 to reduce autocorrelation among parameter samples. For all three imputation methods we select imputation size $M = 50$, and we utilize bootstrap size $B = 100$ for both the proposed wild bootstrap and the nonparametric Bayesian bootstrap given in Gao et al.¹⁶ We utilize the $Exp(1)$ distribution for the wild bootstrap weights, u_i . The results of these analyses and their computational times are presented in Table 4.

Regardless of imputation scenario, all methods result in a treatment effect estimate indicating a mitigating effect of CGM on rate of severe hypoglycemic event recurrence that is not statistically significant. Under CR and J2R assumptions, DI and MLMI point estimates for δ are more similar to each other than to those produced by the Bayesian posterior draw MI method. This likely stems from differences in ML and Bayesian estimation between these methods, but could also indicate poor fit of the Bayesian model. As such, caution is necessary for interpretation of observed trends in model parameter estimation via the posterior draw MI and nonparametric bootstrap.

Under CR and J2R imputation, the proposed DI and wild bootstrap produce the smallest standard errors for $\hat{\delta}$, with the reduction in standard error compared to the other two methods being largest in the case of CR imputation. The posterior draw MI and nonparametric bootstrap of Gao et al¹⁶ produces the largest standard errors for the treatment effect estimator under these control-based assumptions. Under MAR imputation, DI with the wild bootstrap yields the largest standard error for $\hat{\delta}$. These results agree with simulations in that DI with the wild bootstrap is less likely to overestimate the variance of the treatment effect estimator than MI with Rubin's combining rules under CR and J2R imputation, but not under MAR imputation. Furthermore, DI with the wild bootstrap also yields decreased standard errors even compared to the Bayesian MI and nonparametric bootstrap under the control-based assumptions in this application.

Though DI with the wild bootstrap is less computationally efficient than MI with Rubin's combining rules, the time to perform the DI analysis is not prohibitive at less than 18 seconds under each imputation scenario. Furthermore, DI with the wild bootstrap is much more computationally efficient than the Bayesian MI and nonparametric bootstrap proposed

TABLE 4 Sensitivity analysis results for the treatment effect in the CGM trial application.

Method	CR			J2R			MAR		
	Estimate (95% CI)	SE	Time (s)	Estimate (95% CI)	SE	Time (s)	Estimate (95% CI)	SE	Time (s)
DI	-0.411 (-0.964, 0.141)	0.282	12.20	-0.365 (-0.984, 0.254)	0.316	15.41	-0.388 (-1.150, 0.373)	0.388	17.10
MI	-0.413 (-1.020, 0.195)	0.310	0.59	-0.367 (-0.988, 0.254)	0.317	0.51	-0.391 (-1.018, 0.237)	0.320	0.48
NP	-0.343 (-0.960, 0.275)	0.315	3094.54	-0.407 (-1.044, 0.230)	0.325	8544.94	-0.388 (-1.034, 0.259)	0.330	8551.25

Note: The analysis is performed under copy reference (CR), jump to reference (J2R), and as-randomized missing at random (MAR) imputation assumptions using distributional imputation with the wild bootstrap (DI), multiple imputation with Rubin's combining rules (MI), and the Bayesian posterior draw multiple imputation with nonparametric bootstrap method (NP) of Gao et al.¹⁶

by Gao et al.¹⁶ which runs at more than 250 times longer in the CR scenario and more than 500 times longer in the J2R and MAR scenarios than DI.

9 | CONCLUSIONS

Missing data are inevitable in longitudinal clinical trials. Primary analyses are typically performed under the assumption that data are MAR, making sensitivity analyses imperative to assess robustness of primary analysis conclusions to this assumption. Control-based imputation provides a conservative strategy to transparently assess specific MNAR assumptions in superiority trials. Though MI is common in the literature for applying control-based assumptions in sensitivity analyses of clinical trials with recurrent event endpoints,^{5,16,21,23} Rubin's combining rules are often biased for the true sampling variability of the treatment effect estimator due to uncongeniality and MNAR data assumptions imposed in the control-based setting.²⁸ Analytical estimators known to improve variance estimation in MI can be impractical due to complex, model-specific formulas. While the nonparametric bootstrap has successfully improved variance estimation in control-based MI of recurrent events,^{5,16} this method is computationally expensive.

We propose DI as an alternative method yielding improvements over standard MI for conducting control-based sensitivity analyses of recurrent events endpoints. DI with the parallel wild bootstrap variance estimation procedure produces asymptotically consistent point and variance estimators for the treatment effect. In simulations and an application to a type I diabetes trial with recurrent events endpoint, DI yielded point estimates of the treatment effect similar to those produced by MI. DI with the wild bootstrap more accurately estimated the true sampling variability of the treatment effect estimator and produced improved power, more precise confidence intervals, and confidence interval coverage rates and type I error rates closer to nominal levels for the treatment effect estimator under the control-based assumptions of CR and J2R imputation than did MI with Rubin's combining rules in simulations. Researchers may, however, need to exercise caution in regard to type I error rates and confidence interval coverage rates for DI under control-based, and particularly CR, imputation when sample size is small to moderate and dropout rates are high. Simulations and the results of the type I diabetes trial application suggest the improvements demonstrated for DI over MI with Rubin's combining rules do not extend to MAR imputation. Utilizing a nonparametric bootstrap for variance estimation with MI did greatly improve standard error estimation under control-based imputation in simulations. However, most improvements observed in simulations when using the nonparametric bootstrap with MI under control-based imputation were comparable to those shown for DI, and MI with the nonparametric bootstrap was computationally more expensive than DI with the proposed wild bootstrap.

In addition to attractive asymptotic properties and some demonstrated comparative improvements in finite-sample performance, DI and the accompanying wild bootstrap offer simplicity and flexibility for the implementation of control-based assumptions. One limitation of DI compared to MI is its increased storage costs. However, this can be partially addressed by using a weighted regression technique to decrease the amount of storage needed. In this article, we considered monotone missingness resulting from dropout or loss to follow-up, but control-based imputation can handle specification of assumptions according to multiple types of intercurrent events.² We leave the extension of DI of recurrent events to handle intermittent missingness or to incorporate multiple reasons for missingness for future research. We also chose to apply DI to impute missing event counts, but DI could readily be used with the gamma frailty imputation model to impute event times for a time to recurrent events analysis. See Tang³⁸ for practical suggestions for imputation strategies of times to recurrent events. Our chosen imputation and analysis models were fully parametric. One further area of work is to implement DI of recurrent events in the nonparametric or semi-parametric setting, for which the imputation model proposed by Diao et al⁵ could be useful.

ACKNOWLEDGEMENTS

We would like to warmly thank Yilong Zhang and Guanghan Frank Liu for the discussions that motivated our work. We would also like to thank Siyi Liu for initial conversations about distributional imputation that further inspired this research. Yang is partially supported by NSF SES grant 2242776 and the NIH grants 1R01AG066883 and 1R01ES031651.

DATA AVAILABILITY STATEMENT

The JDRF Continuous Glucose Monitoring (CGM) Randomized Clinical Trial data used in this article are openly available in the JAEB Center for Health Research repository at <https://public.jaeb.org/datasets/diabetes>.

ORCID

Sarah R. Fairfax  <https://orcid.org/0009-0000-0124-2165>
Shu Yang  <https://orcid.org/0000-0001-7703-707X>

REFERENCES

1. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592.
2. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide. *Stat Med*. 2020;39(21):2815-2842. doi:[10.1002/sim.8569](https://doi.org/10.1002/sim.8569)
3. ICH. E9(R1) statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials. *FDA Guidance Doc*. 2021. <https://www.fda.gov/media/148473/download>
4. 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):1352-1371. doi:[10.1080/10543406.2013.834911](https://doi.org/10.1080/10543406.2013.834911)
5. Diao G, Liu GF, Zeng D, et al. Efficient multiple imputation for sensitivity analysis of recurrent events data with informative censoring. *Stat Biopharm Res*. 2022;14(2):153-161. doi:[10.1080/19466315.2020.1819403](https://doi.org/10.1080/19466315.2020.1819403)
6. Yang S, Zhang Y, Liu GF, Guan Q. SMIM: a unified framework of survival sensitivity analysis using multiple imputation and martingale. *Biometrics*. 2023;79:230-240. doi:[10.1111/biom.13555](https://doi.org/10.1111/biom.13555)
7. Little RJA. Pattern-mixture models for multivariate incomplete data. *J Am Stat Assoc*. 1993;88(421):125-134. doi:[10.2307/2290705](https://doi.org/10.2307/2290705)
8. Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharm Stat*. 2013;12(6):337-347. doi:[10.1002/pst.1549](https://doi.org/10.1002/pst.1549)
9. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc; 1987.
10. Lu K. An analytic method for the placebo-based pattern-mixture model. *Stat Med*. 2014;33(7):1134-1145. doi:[10.1002/sim.6008](https://doi.org/10.1002/sim.6008)
11. Liu GF, Pang L. On analysis of longitudinal clinical trials with missing data using reference-based imputation. *J Biopharm Stat*. 2016;26(5):924-936. doi:[10.1080/10543406.2015.1094810](https://doi.org/10.1080/10543406.2015.1094810)
12. Tang Y. On the multiple imputation variance estimator for control-based and delta-adjusted pattern mixture models. *Biometrics*. 2017;73(4):1379-1387. doi:[10.1111/biom.12702](https://doi.org/10.1111/biom.12702)
13. Robins JM, Wang N. Inference for imputation estimators. *Biometrika*. 2000;87(1):113-124. doi:[10.1093/biomet/87.1.113](https://doi.org/10.1093/biomet/87.1.113)
14. Yang S, Kim JK. A note on multiple imputation for method of moments estimation. *Biometrika*. 2016;103:244-251. doi:[10.1093/biomet/asv073](https://doi.org/10.1093/biomet/asv073)
15. Guan Q, Yang S. A unified inference framework for multiple imputation using martingales. *Stat Sin*. doi:[10.5705/ss.202021.0404](https://doi.org/10.5705/ss.202021.0404)
16. Gao F, Liu GF, Zeng D, et al. Control-based imputation for sensitivity analyses in informative censoring for recurrent event data. *Pharm Stat*. 2017;16(6):424-432. doi:[10.1002/pst.1821](https://doi.org/10.1002/pst.1821)
17. von Hippel PT, Bartlett JW. Maximum likelihood multiple imputation: faster imputations and consistent standard errors without posterior draws. *Stat Sci*. 2021;36(3):400-420. doi:[10.1214/20-STS793](https://doi.org/10.1214/20-STS793)
18. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol*. 2015;44(1):324-333. doi:[10.1093/ije/dyu222](https://doi.org/10.1093/ije/dyu222)
19. Gill R, Andersen P. Cox's regression model for counting processes: a large sample study. *Ann Stat*. 1982;10(4):1100-1120.
20. Sharef E, Strawderman RL, Ruppert D, Cowen M, Halasyamani L. Bayesian adaptive b-spline estimation in proportional hazards frailty models. *Electron J Stat*. 2010;4:606-642. doi:[10.1214/10-EJS566](https://doi.org/10.1214/10-EJS566)
21. Akacha M, Ogundimu EO. Sensitivity analyses for partially observed recurrent event data. *Pharm Stat*. 2016;15(1):4-14. doi:[10.1002/pst.1720](https://doi.org/10.1002/pst.1720)
22. JDRF CGM Study Group. JDRF randomized clinical trial to assess the efficacy of real-time continuous glucose monitoring in the management of type 1 diabetes: research design and methods. *Diabet Technol Therape*. 2008;10(4):310-321. doi:[10.1089/dia.2007.0302](https://doi.org/10.1089/dia.2007.0302)
23. Keene ON, Roger JH, Hartley BF, Kenward MG. Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharm Stat*. 2014;13(4):258-264. doi:[10.1002/pst.1624](https://doi.org/10.1002/pst.1624)
24. Mitroiu M, Teerenstra S, Rengerink KO, Pétavy F, Roes KC. Estimation of treatment effects in short-term depression studies. An evaluation based on the ICH E9(R1) estimands framework. *Pharm Stat*. 2022;21(5):1037-1057. doi:[10.1002/pst.2214](https://doi.org/10.1002/pst.2214)
25. Wang N, Robins JM. Large-sample theory for parametric multiple imputation procedures. *Biometrika*. 1998;85(4):935-948. doi:[10.1093/biomet/85.4.935](https://doi.org/10.1093/biomet/85.4.935)
26. Bartlett JW. Reference-based multiple imputation—what is the right variance and how to estimate it. *Stat Biopharm Res*. 2021;15(1):178-186. doi:[10.1080/19466315.2021.1983455](https://doi.org/10.1080/19466315.2021.1983455)
27. Yamaguchi Y, Yoshida S, Misumi T, Maruo K. Multiple imputation for longitudinal data using Bayesian lasso imputation model. *Stat Med*. 2022;41(6):1042-1058. doi:[10.1002/sim.9315](https://doi.org/10.1002/sim.9315)
28. Bartlett JW, Hughes RA. Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Stat Methods Med Res*. 2020;29(12):3533-3546. doi:[10.1177/0962280220932189](https://doi.org/10.1177/0962280220932189)
29. Xie X, Meng XL. Dissecting multiple imputation from a multi-phase inference perspective: what happens when God's, imputer's and analyst's models are uncongenial? *Stat Sin*. 2017;27:1485-1594. doi:[10.5705/ss.2014.067](https://doi.org/10.5705/ss.2014.067)
30. Meng XL. Multiple-imputation inferences with uncongenial sources of input. *Stat Sci*. 1994;9(4):538-558. doi:[10.1214/ss/1177010269](https://doi.org/10.1214/ss/1177010269)
31. Cro S, Carpenter JR, Kenward MG. Information-anchored sensitivity analysis: theory and application. *J R Statist Soc A*. 2019;182(2):623-645.

32. Liu S, Yang S, Zhang Y, Liu G. Sensitivity analysis in longitudinal clinical trials via distributional imputation. *Stat Methods Med Res.* 2023;32(2):181-194. doi:[10.1177/09622802221135251](https://doi.org/10.1177/09622802221135251)
33. Lepage GP. A new algorithm for adaptive multidimensional integration. *J Comput Phys.* 1978;27(2):192-203. doi:[10.1016/0021-9991\(78\)90004-9](https://doi.org/10.1016/0021-9991(78)90004-9)
34. Kim JK. Parametric fractional imputation for missing data analysis. *Biometrika.* 2011;98(1):119-132. doi:[10.1093/biomet/asq073](https://doi.org/10.1093/biomet/asq073)
35. Yang S, Kim JK. Fractional imputation in survey sampling: A comparative review. *Stat Sci.* 2016;31(3):415-432. doi:[10.1214/16-STS569](https://doi.org/10.1214/16-STS569)
36. Mammen E. Bootstrap and wild bootstrap for high dimensional linear models. *Ann Stat.* 1993;21(1):255-285.
37. Geweke J. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica.* 1989;57(6):1317-1339. doi:[10.2307/1913710](https://doi.org/10.2307/1913710)
38. Tang Y. Algorithms for imputing partially observed recurrent events with applications to multiple imputation in pattern mixture models. *J Biopharm Stat.* 2018;28(3):518-533. doi:[10.1080/10543406.2017.1333999](https://doi.org/10.1080/10543406.2017.1333999)

SUPPORTING INFORMATION

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

How to cite this article: Fairfax SR, Yang S. Distributional imputation for the analysis of censored recurrent events. *Statistics in Medicine.* 2024;43(13):2622-2640. doi: 10.1002/sim.10087