Received: 29 January 2019 | Revised: 21 July 2019 | Accepted: 12 September 2019

DOI: 10.1002/sim.8385

RESEARCH ARTICLE

WILEY Statistics in Medicine

# Regression analysis of recurrent-event-free time from multiple follow-up windows

**Meng Xia**[1] | **Susan Murray**[1] | **Nabihah Tayob**[2]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan

[2]Department of Data Sciences, Dana-Farber Cancer Institute, Boston, Massachusetts

**Correspondence**
Susan Murray, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109.
Email: skmurray@umich.edu

This research develops multivariable restricted time models appropriate for analysis of recurrent events data, where data is repurposed into censored longitudinal time-to-first-event outcomes in $\tau$-length follow-up windows. We develop two approaches for addressing the censored nature of the outcomes: a pseudo-observation (PO) approach and a multiple-imputation (MI) approach. Each of these approaches allows for complete data methods, such as generalized estimating equations, to be used for the analysis of the newly constructed correlated outcomes. Through simulation, this manuscript assesses the performance of the proposed PO and MI methods. Both PO and MI approaches show attractive results with either correlated or independent gap times in an individual. We also demonstrate how to apply the proposed methods in the data from azithromycin in Chronic Obstructive Pulmonary Disease Trial.

**KEYWORDS**

generalized estimating equation, multiple imputations, multivariable regression, pseudo-observations, recurrent events

## 1 | INTRODUCTION

Recurrent events are frequently seen in participants of clinical trials and observational studies of chronic diseases. For instance, patients in the azithromycin in Chronic Obstructive Pulmonary Disease (COPD) Trial[1] were followed for recurrent acute pulmonary exacerbations. Other settings with recurrent events include recurrent ischemic cardiovascular events after acute coronary syndrome,[2] recurrent clostridium difficile infection,[3] and even repetitive head injuries in high-contact sports.[4] Poisson and negative binomial count models have been used to analyze recurrent event data per time at risk.[5-8] These approaches to not take advantage of the timing of events, however, and may therefore not provide the most powerful analysis.[9]

The most commonly used multivariable regression analysis methods for recurrent events data are extensions of the Cox proportional hazards model to the recurrent event setting. The extension proposed by Anderson and Gill[10] analyzes the time between recurrent events, called gap times, assuming independence between these gap times within an individual. Prentice et al[11] considered an extension of the Cox model that allowed stratification of the baseline hazard to depend on time-dependent features including previous recurrent event time information; both gap time models and models of time from beginning of follow-up are considered. Wei et al[12] proposed a multivariate proportional hazards model, where the multivariate outcomes are based on separate recurrent events modeled from the beginning of follow-up. An arbitrary covariance structure is allowed between the different event times, fit with a robust sandwich variance estimate. Pepe and Cai[13] described several manners of modeling recurrent event rates based on the number of previous recurrent events, advocating for a Markov approach that models each recurrent event conditional on information from the immediately preceding event. Lawless and Nadeau[14] and Lin et al[15] developed models for the cumulative mean number of events,

  

assuming proportionality on the cumulative means over time. Random effects or frailties have been suggested for capturing dependence between recurrent event times.[16-23] Methodology for the analysis of recurrent events continues to evolve via gap time, intensity process or cumulative mean event approaches; a more thorough review is available in a textbook by Cook and Lawless.[24]

In pursuing any new modeling framework for recurrent events, three issues are paramount to address (1) the potential correlation between times between recurrent events, (2) the potentially censored nature of the data, and (3) the interpretability of results. In addressing each of these issues in this manuscript, we take an entirely different approach to modeling recurrent event data that provides a natural way to handle correlation between event times and is highly interpretable. In short, we transform the recurrent event data structure into a very tractable censored longitudinal data structure. The longitudinal outcomes are $\tau$-restricted times-to-first-event as captured in follow-up windows that are reinitiated at regularly spaced intervals. Instead of modeling the rate or cumulative number of recurrent events, our model estimates time free from recurrence over a $\tau$-length follow-up period.

In Section 2, we describe notation required to repurpose traditional recurrent events data into a series of censored longitudinal endpoints. Section 3 describes differences between this data structure and that of the multivariate distribution of gap times between recurrent events. In Section 4, we develop a model framework that can be fit using generalized estimating equation (GEE) methods, along with two methods for handling the censored nature of the data: a pseudo-observation (PO) approach (Section 4.1) and a multiple-imputation (MI) approach (Section 4.2). Section 5 describes finite sample properties of our methodology in scenarios where times between recurrent events are independent (Section 5.1) and correlated (Section 5.2). We then reanalyze data from the azithromycin in COPD Trial using our methodology in Section 6. Discussion follows in Section 7.

## 2 | NOTATION

Suppose $i = 1, \ldots, N$ independent patients are followed for recurrent events. Without loss of generality, we assume each patient's follow-up period starts from a baseline time of 0; hereafter, we refer to baseline and time 0 interchangeably. For each individual patient, $i$, let $T_{ij}, j = 1, \ldots, J_i$, be the time from baseline to the $j^{th}$ recurrent event, so that $0 < T_{i1} < T_{i2} < \cdots T_{iJ_i}$. Let $C_i$ be the censoring time from baseline for patient $i$, where $C_i$ is independent of $T_{ij}$, for $j = 1, \ldots, J_i$. Correlation between recurrent event times in an individual $i$ (or lack thereof) is typically formulated in terms of gap times between events, $\{G_{i1} = T_{i1}, G_{i2} = T_{i2} - T_{i1}, \ldots, G_{iJ_i} = T_{iJ_i} - T_{iJ_{i-1}}\}$. We allow an arbitrary multivariate distribution for $\{G_{i1}, G_{i2}, \ldots, G_{iJ_i}\}$ with an unspecified dependence structure. Traditional observed recurrent event data for patient $i$ is recorded in data pairs $\{X_{ij} = \min(T_{ij}, C_i), \delta_{ij} = I(T_{ij} \leq C_i)\}$, $j = 1, \ldots, \tilde{J}_i$, where $\tilde{J}_i \leq J_i$; in most cases, the $\tilde{J}_i^{th}$ data pair corresponds to a censored event time.

In this manuscript, we construct a streamlined censored longitudinal data structure from the recurrent event times. That is, each longitudinally measured outcome contributed by patient $i$ is a censored time-to-first-event in a follow-up window starting at time $t$, where $t \in \{t_1, \ldots, t_b\}$ with $t_1 = 0$ and $t_k = t_{k-1} + a, k = 2, \ldots, b$. As only one time-to-first-event in each follow-up window is measured, we incorporate at most $b$ outcomes from each individual, regardless of how many recurrent events they experience. Hence, for a fixed overall study duration, the choice of spacing, $a = t_k - t_{k-1}, k = 2, \ldots, b$, between initiation of each subsequent follow-up window increases the proportion of recurrent events captured by the censored longitudinal data structure. It is theoretically possible to create a censored longitudinal dataset with follow-up windows initiated every day ($a = 1$), although the computational burden of working with this extended dataset becomes cumbersome. Xia and Murray[25] showed that, in the case of exponentially distributed times between events with common intensity $\lambda$, using $a = 1/(2\lambda)$ captures approximately 80% of the recurrent events in at least one of the constructed follow-up windows over a fixed follow-up period.

For patient $i$ and follow-up window starting at $t$, we index the first recurrent event occurring after time $t$ with the subscript $\eta_i(t) = \min\{j = 1, \ldots, J_i : T_{ij} \geq t\}$ so that $T_i(t) = T_{i\eta_i(t)} - t$ is the time-to-first-recurrent-event measured from $t$, sometimes called the residual event-free time from $t$. We collect individual $i$'s newly formatted longitudinal outcomes, $\{T_i(t_1), T_i(t_2), \ldots, T_i(t_b)\}$, into a vector, $\mathcal{T}_i$, for $i = 1, \ldots, N$. For notational simplicity, we occasionally submerge the individual $i$ index when it is not required for clarity. For example, we will often use $\Pr\{T(t) > u\}$ to stand in for $\Pr\{T_i(t) > u\}$.

The observed data counterpart to $\eta_i(t)$ is $\tilde{\eta}_i(t) = \min\{j = 1, \ldots, \tilde{J}_i : X_{ij} \geq t\}$. For each follow-up window starting at time $t$ where $C_i > t$, patient $i$ contributes the observed data triplet $\{\tilde{\eta}_i(t), X_i(t) = X_{i\tilde{\eta}_i(t)} - t, \delta_i(t) = \delta_{i\tilde{\eta}_i(t)}\}$. For follow-up windows starting at $t$ where $C_i \leq t$, we use the convention that $\tilde{\eta}_i(t) = X_i(t) = \delta_i(t) = 0$.

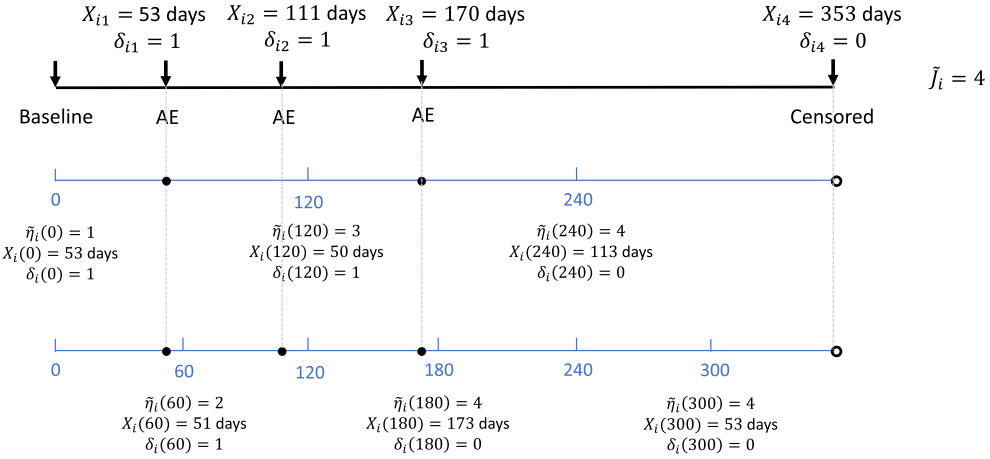


**FIGURE 1** Creating censored longitudinal data from recurrent event data for an example participant of the azithromycin in Chronic Obstructive Pulmonary Disease Trial. AE, acute exacerbation [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 1 displays how censored longitudinal data is created from traditional recurrent event data using a participant from the azithromycin in COPD Trial. During 353 days of follow-up for this patient, $\tilde{J}_i = 4$ traditional recurrent event data pairs emerge: $(X_{i1} = 53 \text{ days}, \delta_{i1} = 1)$, $(X_{i2} = 111 \text{ days}, \delta_{i2} = 1)$, $(X_{i3} = 170 \text{ days}, \delta_{i3} = 1)$, and $(X_{i4} = 353 \text{ days}, \delta_{i4} = 0)$, where the first three data pairs denote acute exacerbation (AE) event times and the last data pair reflects a censored event.

Two examples of converting these data into a censored longitudinal data structure are given, one based on follow-up windows starting at $a = 120$ day intervals and one constructed with $a = 60$ day intervals. As with all longitudinal data structures, the additional data triplets included using $a = 60$ day intervals as opposed to $a = 120$ day intervals afford capturing more time-to-first events supplied by the recurrent event times.

Data triplets based on follow-up windows starting at $t = \{0, 120, 240\}$ days become

$$\{\tilde{\eta}_i(0) = 1, X_i(0) = X_{i1} - 0 = 53, \delta_i(0) = 1\},$$
$$\{\tilde{\eta}_i(120) = 3, X_i(120) = X_{i3} - 120 = 170 - 120 = 50, \delta_i(120) = 1\},$$
$$\{\tilde{\eta}_i(240) = 4, X_i(240) = X_{i4} - 240 = 353 - 240 = 113, \delta_i(240) = 0\}.$$

Data triplets based on follow-up windows initiated every $a = 60$ days include the above data triplets plus those starting at days $t = \{60, 180, 300\}$ days

$$\{\tilde{\eta}_i(60) = 2, X_i(60) = X_{i2} - 60 = 111 - 60 = 51, \delta_i(60) = 1\},$$
$$\{\tilde{\eta}_i(180) = 4, X_i(180) = X_{i4} - 180 = 353 - 180 = 173, \delta_i(180) = 0\},$$
$$\{\tilde{\eta}_i(300) = 4, X_i(300) = X_{i4} - 300 = 353 - 300 = 53, \delta_i(300) = 0\}.$$

## 3 | TIMES-TO-FIRST-EVENT FROM $t \in \{t_1, \ldots, t_b\}$ VERSUS GAP TIMES

The marginal distribution of the time-to-first recurrent event after $t$, $T_i(t)$, is a quite different creature from the marginal distribution of a gap time between recurrent events, $G_{ij}, j = 1, \ldots, J_i$. On a practical note, the random variable, $T_i(t)$, better reflects the recurrent event time one might seek advice about at a regularly scheduled clinic visit at time $t$ or at entry into a clinical trial at $t$. These patient interactions rarely coincide exactly with a recurrent event, so that a gap time random variable measured from an individual's previous event is not the most appropriate random variable for these settings. On a statistical note, when gap times within an individual are correlated there is a well-known dependent censoring bias that must be addressed in any analysis of the gap time data.[26] This dependent censoring issue is circumvented by our censored longitudinal data structure since times-to-first-event are measured from pre-specified times $\{t_1, \ldots, t_b\}$ rather than a correlated time-to-event.

For these different random variables to coincide with one another, the distributions of $T_i(t)$ and $G_{ij}$ must both be entirely memoryless. This is formally demonstrated in the following, where we show that the only case where the marginal distribution of a gap time, $G_{ij}$, coincides with that of a time-to-first-recurrent-event, $T_i(t)$, is the special case with independent and identically distributed (i.i.d.) exponential gap times $\{G_{i1}, \ldots, G_{iJ_i}\}$ for each patient, $i = 1, \ldots, N$. For settings with gap times that are not exponentially distributed, or for settings with correlated, but otherwise identically distributed,

exponential gap times, the marginal distribution of $T_i(t)$ shifts from a memoryless distribution to a distribution very much influenced by the series of recurrent events with positive probabilistic support for occurring in the follow-up period after $t$.

Consider the event-free probability function for $T_i(t)$ that is written in terms of gap time random variables as follows:

$$\begin{aligned}\Pr\{T_i(t) > u\} &= \Pr\{T_{i1} > t+u\} + \lim_{J_i\to\infty}\sum_{j=2}^{J_i}\Pr\{T_{ij-1} \le t, T_{ij} > t+u\} \\ &= \Pr\{G_{i1} > t+u\} + \lim_{J_i\to\infty}\sum_{j=2}^{J_i}\Pr\left\{\sum_{l=1}^{j-1}G_{il} \le t, \sum_{l=1}^{j-1}G_{il} + G_{ij} > t+u\right\}. \end{aligned} \tag{1}$$

For independently and identically distributed exponential gap times with intensity $\lambda$, Appendix A of the Supplementary Materials shows that these terms reduce to $\exp(-\lambda u)$, so that $T_i(t)$ also has an exponential distribution with intensity $\lambda$.

However, when the gap times are correlated, the term

$$\Pr\left\{\sum_{l=1}^{j-1}G_{il} \le t, \sum_{l=1}^{j-1}G_{il} + G_{ij} > t+u\right\} \tag{2}$$

from the previous equation does not reduce to a simple expression. Term (2) is the probability that an individual's $j^{th}$ recurrent event will be the first to occur in the follow-up window starting at $t$, but that it has not yet occurred as of time $t+u$.

To better appreciate the influence of term (2) on the expression in (1), we consider special cases with correlated and independent exponential($\lambda_i$) distributed times between recurrent events. Figure S3 in the Supplementary Materials displays term (2) as a function of time, $u$, for different combinations of recurrent event index, $j$, and follow-up window start time, $t$, with $\lambda_i = 1/3$. The solid blue and dashed red lines show the cases with independent and correlated event times, respectively. For the independence case, the curves have a closed-form shown to be $(\lambda_i t)^{j-1}e^{-\lambda_i(t+u)}/\Gamma(j)$ in Appendix A of the Supplementary Materials. For the correlated case, we first simulated correlated exponential event times using a Gaussian copula approach described in further detail in Section 5; the approximate correlation between recurrent event times was 0.8. We then empirically estimated and plotted term (2) from a large number ($N = 10\,000$) of simulated individuals.

In nearly every panel of Figure S3, term (2) is smaller when times between recurrent events are correlated. As $j$ increases relative to $t$ the depicted probability curves get lower, and the curves generated from the two different correlation structures also get closer together. Overall, these curves indicate that $\Pr\{T_i(t) > u\}$ tends to be much smaller than the exponential(1/3) survival curve that Equation (1) reduces to in the case with independent event times.

Of course, we were immediately curious to know whether the distribution of the time-to-first-event from $t$, in this special case with correlated exponential(1/3) gap times, stabilizes. The intuition behind this thought was that the mixture distribution of gap time histories preceding $t$ and likely to influence the distribution of $T_i(t)$ might stabilize. As seen in Figure S4, where (again using the large simulated dataset of 10 000 individuals) $\Pr\{T_i(t) > u\}$ is plotted for increasing values of $t$, this does seem to be the case. The distribution of $T_i(t)$ seems to stabilize for values of $t \ge 3$, or $1/\lambda_i$. Stabilization of the distribution of $T_i(t)$, for $t > 1/\lambda_i$ was further explored for different values of $\lambda_i$ and found to be a reliable pattern. This feature will be utilized later in Section 5.2, when we simulate a stable time-to-first-event distribution given covariates in the setting with correlated exponential gap times.

## 4 | MULTIVARIABLE REGRESSION MODEL OF $\tau$-RESTRICTED TIMES-TO-FIRST-RECURRENT-EVENT MEASURED ACROSS MULTIPLE OVERLAPPING FOLLOW-UP PERIODS

To study the association between patient covariates, $Z(t)$, and $\tau$-restricted times-to-first-recurrent event across follow-up windows starting at times $\{t_1, \ldots, t_b\}$, we consider the following model:

$$E\{\log[\min(\tau, \mathcal{T})]|Z(t)\} = \beta^T Z(t). \tag{3}$$

For simplicity in what follows, we submerge the time-dependent nature of the covariates unless needed for clarity. As with standard $\tau$-restricted mean time-to-event models, $\tau$ is a user-defined parameter that is chosen to reflect a time-period

of clinical interest. For instance $\tau = 6$ months in model (3) reflects interest in estimating the typical event-free period that an individual can expect over the next 6 months. For each follow-up window starting at $t_k, k = 1, \ldots, b$, the $\tau$-restricted event time for individual $i$ is $\min[\tau, X_i(t_k)]$ with censoring indicator $\max(\delta_i(t_k), I\{X_i(t_k) \geq \tau\})$. If individual $i$ does not experience an event within a particular observed follow-up window starting at $t_k$, the observed $\tau$-restricted event-free time is $\tau$.

Two features of our data need to be addressed for successful estimation of model (3): (1) the censored nature of the vector of newly formatted longitudinal outcomes, $\mathcal{T}_i$, from patient $i$, and (2) the correlated nature of these longitudinal outcomes. We develop two approaches that address the censoring aspect of the data, a PO approach in Section 4.1 and a MI approach in Section 4.2. Each of these approaches converts the censored longitudinal outcomes into a format appropriate for complete data methods.

Once this feature of the data is addressed, we tackle the correlated nature of the longitudinal outcomes $\mathcal{T}_i$ from each patient, $i = 1, \ldots, N$ using existing methods, such as GEEs. Because our recurrent events data have been restructured into times-to-first event from regularly spaced follow-up periods and because we consider $\tau$-restricted times-to-first-events in these periods, the correlation structure of the outcomes can be modeled via well-organized correlation matrices.

Two underlying layers of correlation are at work: the natural correlation between recurrent events within an individual and the possibility that the same event is captured as the first-time-to-event in more than one follow-up period. In the most general case, we assume an $b \times b$ unstructured correlation matrix with components

$$\begin{bmatrix} 1 & \text{corr}\{T_i(t_1), T_i(t_2)\} & \text{corr}\{T_i(t_1), T_i(t_3)\} & \cdots & \text{corr}\{T_i(t_1), T_i(t_b)\} \\ \text{corr}\{T_i(t_2), T_i(t_1)\} & 1 & \text{corr}\{T_i(t_2), T_i(t_3)\} & \cdots & \text{corr}\{T_i(t_2), T_i(t_b)\} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \text{corr}\{T_i(t_b), T_i(t_1)\} & \text{corr}\{T_i(t_b), T_i(t_2)\} & \text{corr}\{T_i(t_b), T_i(t_3)\} & \cdots & 1 \end{bmatrix}.$$

Of course, pre-specified follow-up windows that begin after a patient's last observed follow-up time do not contribute any information for analysis, resulting in variable cluster sizes per subject in practice. This does not pose any difficulty for the GEE method.[27,28]

However, for settings with fairly stable time-to-first-event distributions over time, we consider a (banded) Toeplitz correlation structure that allows for correlation to decrease as the degree of overlap between $\tau$-restricted follow-up periods decreases. The degree of overlap in $\tau$-restricted follow-up windows can be determined from $a$ and $\tau$ and follows a regular pattern. For instance with $a = \tau/3$ and $b = 4$ windows starting at times $t = \{0, \tau/3, 2\tau/3, \tau\}$, the Toeplitz correlation matrix is

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix},$$

where $\rho_1$ is the correlation between times-to-first event in adjacent $\tau$-restricted follow-up windows that overlap by $2\tau/3$ follow-up units, $\rho_2$ is the correlation between times-to-first event in windows that start $2a$ units apart from one another and overlap by $\tau/3$ units. Finally, $\tau$-restricted follow-up windows starting $a = 3$ units apart from one another do not overlap and are assumed to have correlation $\rho_3$.

The Toeplitz correlation structure requires fewer parameters than the unstructured matrix. For very large $b$, the Toeplitz correlation structure may be more feasible to implement that an entirely unstructured variance matrix. GEE also provides model results based on robust sandwich variance estimation, which provides protection against misspecification of the working correlation matrix. The general recommendation when working with large datasets is to use the sandwich estimator, regardless of the working correlation matrix assumed by the model. We follow this recommendation throughout the remainder of the manuscript.

## 4.1 | PO approach for censored recurrent events

For a single time-to-event, Andersen et al[29] introduced the idea of using POs in lieu of censored times-to-event when estimating regression parameters for the restricted mean model. This method has been successfully applied in a variety of settings where a single event time is of interest.[30-35] The appeal of this method is its ease of use. That is, once appropriate POs are estimated for each patient, they can be used as if they are uncensored counterparts to the original censored

data in standard regression models. In this section, we describe how to create POs that correspond to our censored longitudinal data structure. In particular, for each follow-up window starting at $t$, we define POs for the random variables $\log[\min\{\tau, T_i(t)\}], i = 1, \ldots, N$, using a method similar to that described by Xiang and Murray.[34]

The general intuition behind PO approaches for modeling censored survival data is similar to that of the jackknife method.[36-38] One first defines a consistent nonparametric estimate, $\hat{\theta}$, of the marginal mean of interest, $\theta$. In our setting, for each $t \in \{t_1, \ldots, t_b\}$, we define $\theta(t) = E[\log\{\min(\tau, T(t))\}]$ with consistent nonparametric estimator

$$\hat{\theta}(t) = -\int_0^{\tau} \log(u) d\hat{P}(T(t) > u) + \log(\tau)\hat{P}(T(t) > \tau),$$

where Kaplan-Meier estimation is used for $\hat{P}(T(t) > u)$.

The form of an appropriate PO for any setting arises from framing $\theta$ both as a marginal mean and a weighted average of $\theta_Z$, the conditional mean given covariates, $Z$. Most readers will recognize this relationship when formally depicted as

$$\theta = \int \theta_Z dF_Z(z),$$

where $dF_Z(z)$ reflects Riemann-Stieltjes integration across the distribution of $Z$. When the empirical (discrete) distribution of $Z$ is used in framing the relationship above, $dF_Z(z) = 1/N$, and the right-hand side of the expression becomes

$$\frac{1}{N}\sum_{i=1}^{N}\theta_{Z_i}.$$

One can algebraically isolate $\theta_{Z_i}$ (individual $i$'s mean given $Z_i$) from the expression above via

$$\theta_{Z_i} = N\left\{\frac{1}{N}\sum_{j=1}^{N}\theta_{Z_j}\right\} - (N-1)\left\{\frac{1}{N-1}\sum_{j=1, j\neq i}^{N}\theta_{Z_j}\right\}.$$

Marginal means corresponding to the terms in curly brackets can be consistently estimated using nonparametric estimates $\hat{\theta}$ and $\hat{\theta}^{(-i)}$, respectively, where $\hat{\theta}^{(-i)}$ is the "leave-one-out" estimator of $\theta$, ie, estimated without individual $i$. So, taking advantage of large sample properties of $\hat{\theta}$ and $\hat{\theta}^{(-i)}$, a natural PO for individual $i$ to use in modeling $\theta_Z$ is

$$N\hat{\theta} - (N-1)\hat{\theta}^{(-i)},$$

a fully observed random variable that asymptotically shares a conditional mean, $\theta_{Z_i}$, with patient $i$.

In our setting, $\theta_Z = E(\log[\min\{\tau, \mathcal{T}\}]|Z)$. For each $t$, we define POs

$$PO_i(t) = N\hat{\theta}(t) - (N-1)\hat{\theta}^{(-i)}(t), i = 1, \ldots, N,$$

where

$$\hat{\theta}^{(-i)}(t) = -\int_0^{\tau} \log(u) d\hat{P}^{(-i)}(T(t) > u) + \log(\tau)\hat{P}^{(-i)}(T(t) > \tau),$$

where leave-one-out Kaplan-Meier estimation is used for $\hat{P}^{(-i)}(T(t) > u)$, ie, excluding patient $i$. We denote the vector of POs contributed by individual $i$ as $PO_i = \{PO_i(t_1), PO_i(t_2), \ldots, PO_i(t_b)\}$. Parameter estimates for model (3) can be estimated using the longitudinally created PO data via

$$E[PO|Z] = \beta^T Z. \tag{4}$$

Hereafter, we refer to estimates from Equation (4) as estimates using the proposed PO approach.

### 4.2 | MI approach for censored recurrent events

Another approach for producing a complete dataset when a single time-to-event is subject to censoring is MI. This approach has been developed by many authors.[35,39-43] For our longitudinal data structure, we propose multiply imputing outcomes for observed data pairs with $\{X_i(t) > 0, \delta_i(t) = 0\}, t \in \{t_1, \ldots, t_b\}$.

For our longitudinal data structure, the $i^{th}$ individual requires imputation for times-to-first-event in the set, $\mathcal{S}_i$, of follow-up windows starting at times $\{t \in \{t_1, \ldots, t_b\} : X_i(t) > 0, \delta_i(t) = 0\}$. If $\mathcal{S}_i$ consists of more than one follow-up window, it suffices to impute the time-to-first-event corresponding to the window starting at follow-up time $t^{sup}(\mathcal{S}_i) = \max\{\text{follow-up window start time } t \text{ for windows} \in \mathcal{S}_i\}$, which then determines imputes for all times-to-first-event in the set of follow-up windows, $\mathcal{S}_i$, that require imputation (see Appendix B of the Supplementary Materials for further details). For better short-hand terminology, we call the imputed event time corresponding to follow-up window start time $t^{sup}(\mathcal{S}_i)$ the "sup impute", denoted as $\tilde{T}_i\{t^{sup}(\mathcal{S}_i)\}$, and the follow-up window starting at time $t^{sup}(\mathcal{S}_i)$ the "sup window". Then, imputed event times for follow-up windows in $\mathcal{S}_i$ with start times $t^* < t^{sup}(\mathcal{S}_i)$ become $\tilde{T}_i\{t^*\} = \tilde{T}_i\{t^{sup}(\mathcal{S}_i)\} + t^{sup}(\mathcal{S}_i) - t^*$.

The gestalt of the imputation strategy is to base the sup impute in the sup window on model (4) using individual $i$'s covariates, $Z_i$. Random error for the sup impute is sampled nonparametrically from a set of residuals contributed by individuals in a risk set, $\mathcal{R}_i$, similar to individual $i$. Further details are described below.

The first step of the imputation procedure is to obtain parameter estimates, $\hat{\beta}^{PO}$, from model (4). For individual $i$ requiring a sup impute, $\tilde{T}_i\{t^{sup}(\mathcal{S}_i)\}$, in the sup window, we then define a risk set, $\mathcal{R}_i$, of candidate individuals $l = 1, \ldots, N_i$ satisfying two constraints: (1) $X_l\{t^{sup}(\mathcal{S}_i)\} > X_i\{t^{sup}(\mathcal{S}_i)\}$, that is, candidate $l$ is still at risk for their first event time in the sup window as of the time individual $i$ is censored and (2) $|\hat{\beta}^{PO^T} Z_i - \hat{\beta}^{PO^T} Z_l| \leq \epsilon$, where $\epsilon$ is a user-defined parameter that controls how similar individual $l$'s linear predictor is to individual $i$'s linear predictor. Our algorithm used $\epsilon = 0.01$. In cases where $\epsilon$ resulted in a risk set with $N_i < 5$, our algorithm added 0.001 to $\epsilon$ until $N_i \geq 5$.

The next step of the imputation procedure is use candidate individuals, $l = 1, \ldots, N_i, \in \mathcal{R}_i$ to estimate the survival function for $T\{t^{sup}(\mathcal{S}_i)\}$ given membership in $\mathcal{R}_i$. Nonparametric Kaplan-Meier estimation is used for this purpose, resulting in estimate, $\hat{S}_{T\{t^{sup}(\mathcal{S}_i)\}}(v|\mathcal{R}_i)$. Then, an inverse transform imputation algorithm[35,40-43] is used to select an impute following the distribution of $T\{t^{sup}(\mathcal{S}_i)\}$ given membership in $\mathcal{R}_i$ based on $\hat{S}_{T\{t^{sup}(\mathcal{S}_i)\}}(v|\mathcal{R}_i)$. In particular, the inverse transform imputation method first generates a uniform(0, 1) random variable, $u$. If $\hat{S}_{T\{t^{sup}(\mathcal{S}_i)\}}(v|\mathcal{R}_i) > u$ for all observed event times $v$, we impute $\tilde{T}_i\{t^{sup}(\mathcal{S}_i)\} = \tau$. Otherwise, we find the smallest value $v$ where $\hat{S}_{T\{t^{sup}(\mathcal{S}_i)\}}(v|\mathcal{R}_i) \leq u$ and identify the observed event time, $T_l\{t^{sup}(\mathcal{S}_i)\}$, that corresponds to $v$.

The inverse transform impute for patient $i$'s time-to-first-event in the sup window would be $T_l\{t^{sup}(\mathcal{S}_i)\}$. However, our proposed imputation algorithm goes one step further, by defining residual $\varepsilon_l = \log(\min[\tau, T_l\{t^{sup}(\mathcal{S}_i)\}]) - \hat{\beta}^{PO^T} Z_l$ and then defining our final impute $\tilde{T}_i\{t^{sup}(\mathcal{S}_i)\} = \exp[\hat{\beta}^{PO^T} Z_i + \varepsilon_l]$. This extra step allows for variability of the impute to be contributed by individual $l$, while further targeting the impute using individual $i$'s covariate structure. If $\tilde{T}_i\{t^{sup}(\mathcal{S}_i)\} < X_i\{t^{sup}(\mathcal{S}_i)\}$, we sample another uniform(0, 1), $u$, and repeat the process. This ensures that the impute occurs beyond the last observed time participant $i$ was at risk for a recurrent event in the sup window.

We repeat the imputation procedure until we obtain $M$ completed datasets and then analyze the $M$ imputed datasets with methods guided by Little and Rubin.[44] For dataset $m$, we fit model (3) using GEE as described at the beginning of this section and obtain parameter estimates, $\hat{\beta}_m^{MI}$, and standard error (SE) estimates, $\hat{SE}(\hat{\beta}_m^{MI})$, for $m = 1, \ldots, M$. Then, the final estimate of $\beta$ from the MI procedure becomes $\hat{\beta}^{MI} = \sum_{m=1}^{M} \hat{\beta}_m^{MI}/M$ with corresponding SE estimate

$$\hat{SE}(\hat{\beta}^{MI}) = \sqrt{\sum_{m=1}^{M} \hat{SE}\left(\hat{\beta}_m^{MI}\right)^2/M + (1 + M^{-1}) \times \sum_{m=1}^{M} \left(\hat{\beta}_m^{MI} - \hat{\beta}^{MI}\right)^2/(M-1)}.$$

## 5 | SIMULATION

We now evaluate the finite sample performance of the proposed PO and MI methods for fitting equation (3) with simulated recurrent event data from $N = 500$ individuals over 5 years of follow-up. All simulation results are based on 10 000 iterations. Details of how recurrent events times are simulated are described in Sections 5.1 and 5.2, where independent and dependent recurrent event distributions are considered, respectively. In each simulation scenario, we build our longitudinal data structure with follow-up windows starting every $a = 1$ years apart at times $t = 0, 1, 2$, and 3 years with $\tau = 2$ years. These choices coincide with recommendations from Xia and Murray[25] based on the recurrent event distributions used in simulation.

We explore scenarios with no censoring, light censoring, and heavy censoring. The uncensored case provides a reference point to gauge the impact of censoring on analyses conducted using the PO and MI approaches. In the uncensored case, after restructuring the recurrent events into our recommended longitudinal data structure, a standard GEE approach is used to fit model (3). When censoring is present in our recommended longitudinal data structure, the standard GEE approach is no longer appropriate for fitting model (3) until further processing of the data is done via the PO or MI methods developed in Sections 4.1 and 4.2.

In scenarios where censoring is present, the independent censoring random variable is $C_i = 5 \times I\{V_i > 5\} + V_i \times I\{V_i \leq 5\}, i = 1, \ldots, 500$. For the light censoring case, $V_i$ has an exponential distribution with hazard 1/14, corresponding to approximately 70% of participants having 5 years of follow-up and 30% of participants being subject to censoring prior to 5 years. In the heavy censoring case, $V_i$ has an exponential distribution with hazard 1/4, corresponding to approximately 70% of participants being subject to censoring prior to 5 years.

With four follow-up windows generating four longitudinal outcomes, we require a $4 \times 4$ working correlation structure to use with GEE software. We consider both (1) unstructured and (2) Toeplitz structures, with robust sandwich estimates ultimately used in all inference. The Toeplitz structure takes the form

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_2 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_2 & \rho_1 & 1 \end{bmatrix}.$$

Since only adjacent $\tau$-restricted follow-up windows have overlap in our setting with $t = 0, 1, 2, 3$ and $\tau = 2$, this structure requires only two parameters, as opposed to six parameters used with the unstructured working correlation matrix.

## 5.1 | Independent times between recurrent events

We first describe the scenario where times between recurrent events are independent. Recall from Section 3 and Appendix A that when gap times between events for individual, $i$, are i.i.d. exponential($\lambda_i$) random variables, then $T_i(t)$ is also marginally distributed as an exponential($\lambda_i$) random variable for each $t \in \{t_1, \ldots, t_b\}$. Hence, if we generate times-to-first-event from i.i.d. exponential gap times, the mean structure for $T_i(t), t \in \{t_1, \ldots, t_b\}$, will follow the same mean structure as the simulated gap times. This offers some computational convenience for generating outcomes that follow model (3).

For the $i^{th}$ individual, we allow $\lambda_i$ to depend on two covariates, ie, $Z_i = \{B_i, U_i\}$, where $B_i$ a Bernoulli(0.5) random variable and $U_i$ is a uniform(0, 1) random variable. We generate mild correlation between $B_i$ and $U_i$ using a Gaussian copula approach.[45] That is, we first generate bivariate normal(0, 1) pairs $(Q_{1i} = q_{1i}, Q_{2i} = q_{2i})$ with correlation 0.3. We then define $B_i = I(Q_{1i} \geq 0)$ and $U_i = P(Q_{2i} \leq q_{2i})$; the uniform(0, 1) distribution of $U_i$ follows from the inverse transform theorem. Finally, we generate times-to-first-event for patient $i$, $\mathcal{T}_i$, that satisfy model

$$E(\log[\min\{\tau, \mathcal{T}\}]|Z) = -0.7 + 0.5B_i + 0.5U_i.$$

This is accomplished by first simulating i.i.d. exponential($\lambda_i$) gap times with $\lambda_i$ taken as the numerical solution to

$$\int_{-\infty}^{\log \tau^-} y\lambda_i e^y e^{-\lambda_i e^y} dy + e^{-\lambda_i \tau} \times \log \tau = \beta_0 + \beta_1 B_i + \beta_2 U_i$$

and then converting the resulting recurrent event times into times-to-first event as described in Section 2.

Simulation results for the case with independent times between recurrent events are shown in Table 1. For each method and each coefficient, we present simulation averages for (1) $\hat{\beta}$, (2) bias $\hat{\beta} - \beta$, and estimated robust SEs assuming (3) unstructured or (4) Toeplitz working correlation matrices. We also report (5) the empirical standard deviation (ESD) of $\hat{\beta}$ across the 10 000 iterations and empirical coverage probabilities (CPs) for the true coefficient using robust SEs and either (6) unstructured or (7) Toeplitz working covariance matrices.

All proposed approaches yield approximately unbiased estimates, with absolute bias $\leq 0.01$. CPs are suitably close to 0.95. SE results are very close to ESD results in all scenarios, indicating that variability is being estimated well across all methods. As expected, SEs increase as the percentage of patients subject to censoring increases. SEs attributed to the MI

**TABLE 1** Simulated finite sample performance for $N = 500$ individuals with independently generated times between recurrent events. Results are based on 10 000 iterates

| Coef. | % Cens. | Method | $\hat{\beta}$ | Bias | ESD | SE Unstr. | SE Toepl. | CP Unstr. | CP Toepl. |
|---|---|---|---|---|---|---|---|---|---|
| $\beta_0 = -0.7$ | 0 | GEE | −0.699 | 0.001 | 0.058 | 0.057 | 0.057 | 0.944 | 0.945 |
| | 30 | PO | −0.701 | −0.001 | 0.062 | 0.061 | 0.061 | 0.944 | 0.944 |
| | 30 | MI | −0.697 | 0.003 | 0.062 | 0.060 | 0.060 | 0.943 | 0.943 |
| | 70 | PO | −0.699 | 0.001 | 0.071 | 0.070 | 0.070 | 0.942 | 0.944 |
| | 70 | MI | −0.690 | 0.010 | 0.070 | 0.068 | 0.069 | 0.938 | 0.939 |
| $\beta_1 = 0.5$ | 0 | GEE | 0.500 | <0.001 | 0.054 | 0.054 | 0.054 | 0.945 | 0.946 |
| | 30 | PO | 0.500 | <0.001 | 0.058 | 0.057 | 0.057 | 0.945 | 0.946 |
| | 30 | MI | 0.499 | −0.001 | 0.058 | 0.057 | 0.057 | 0.944 | 0.945 |
| | 70 | PO | 0.498 | −0.002 | 0.067 | 0.066 | 0.066 | 0.945 | 0.945 |
| | 70 | MI | 0.496 | −0.004 | 0.067 | 0.064 | 0.065 | 0.937 | 0.937 |
| $\beta_2 = 0.5$ | 0 | GEE | 0.499 | −0.001 | 0.093 | 0.092 | 0.092 | 0.945 | 0.945 |
| | 30 | PO | 0.499 | −0.001 | 0.100 | 0.098 | 0.099 | 0.945 | 0.945 |
| | 30 | MI | 0.499 | −0.001 | 0.099 | 0.097 | 0.098 | 0.943 | 0.944 |
| | 70 | PO | 0.497 | −0.003 | 0.114 | 0.113 | 0.113 | 0.947 | 0.948 |
| | 70 | MI | 0.497 | −0.003 | 0.114 | 0.111 | 0.111 | 0.941 | 0.942 |

(Coef., true value of the coefficient; % Cens., percent of individuals subject to censoring prior to 5 years of follow-up. For methods: GEE, standard generalized estimating equation approach applied to uncensored version of the data; PO, pseudo-observation approach; MI, multiple-imputation approach. For remaining column headings: $\hat{\beta}$, average coefficient estimate; Bias, average $\hat{\beta} - \beta$; ESD, empirical standard deviation of $\hat{\beta}$; SE Unstr., the average estimated robust standard error using an unstructured working correlation matrix; SE Toepl., the average estimated robust standard error using a Toeplitz working correlation matrix; CP Unstr., empirical coverage probability for true coefficient based on 95% confidence interval using robust standard error with an unstructured working correlation matrix; CP Toepl., empirical coverage probability for true coefficient based on 95% confidence interval using robust standard error with an Toeplitz working correlation matrix.)

method are negligibly smaller than those using the PO method. Both proposed PO and MI analysis methods perform well in settings with either light or heavy censoring. In practice, the PO method is particularly easy to program compared to the MI method and runs a bit more quickly, since the PO method is nested within the MI method. We suspect the PO method will be implemented more in practice as a result.

Additional simulation results with i.i.d. Weibull gap times are given in Appendix C of the Supplementary Materials, with window start times $\{t_1, \ldots, t_b\}$ treated as time-dependent covariates. Operating characteristics are similarly satisfactory in this setting.

## 5.2 | Simulating distribution of times-to-first-event based on correlated times between recurrent events and comparison of proposed methods

Simulating a multivariate time-to-first-event distribution is more complex when times between events are correlated random variables. Recall from Section 3 that positive correlation between exponential($\lambda_i$) gap times causes the corresponding distribution for $\mathcal{T}$ to change; that $P\{T_i(t) > u\}$ tends to be smaller than an exponential($\lambda_i$) survival function and stabilizes after approximately $t > 1/\lambda_i$ follow-up units of gap-time history has passed. The intuition behind this phenomenon, described in Section 3, also suggests the approach for successfully simulating the desired stable multivariable distribution for $\mathcal{T}$ to be used in this section. That is, upon simulating correlated exponential($\lambda_i$) gap times for individual $i$, we discard at least the first $1/\lambda_i$ follow-up units of generated information, starting $t_1 = 0$ for individual $i$ after this "burn-in" period has passed.

To verify that our model works correctly for finite sample sizes when exponentially distributed times between event are correlated, we need (1) to generate data that follows model (3) for this setting and (2) have a way to verify that estimated parameters appropriately represent the data. To address (2), we assume a categorical predictor, $Z_i = 0, 1, 2$, so that $E(\log[\min\{\tau, T_i(t)\}]|Z)$ can be consistently estimated from a large dataset ($N = 10\,000$) within each level of $Z$ via an empirical mean. From this model-free process, we can determine values, $\tilde{\beta}$, of regression parameters that should be estimated if model (3) is working correctly. In particular, we assume that individuals with $Z_i = 0$, 1, or 2 have a history of exponential gap times with $\lambda_i = 1/2, 1/3$ or $1/5$, respectively, where correlation between any two gap times from individual $i$ is approximately 0.8. Stabilization of the resulting multivariate time-to-first-event process is done by defining $t_1 = 0$ after a burn-in period of 5 follow-up units has passed for each individual.

**TABLE 2** Simulated finite sample performance for $N = 500$ individuals with correlated times between recurrent events. Results are based on 10 000 iterates

| Coef. | % Cens. | Method | $\hat{\beta}$ | Bias | ESD | SE Unstr. | SE Toepl. | CP Unstr. | CP Toepl. |
|---|---|---|---|---|---|---|---|---|---|
| $\tilde{\beta}_0 = -0.677$ | 0 | GEE | −0.669 | 0.008 | 0.076 | 0.075 | 0.075 | 0.941 | 0.940 |
| | 30 | PO | −0.667 | 0.010 | 0.079 | 0.078 | 0.078 | 0.939 | 0.939 |
| | 30 | MI | −0.669 | 0.008 | 0.079 | 0.077 | 0.077 | 0.937 | 0.939 |
| | 70 | PO | −0.662 | 0.015 | 0.085 | 0.083 | 0.084 | 0.933 | 0.936 |
| | 70 | MI | −0.668 | 0.009 | 0.085 | 0.082 | 0.083 | 0.935 | 0.936 |
| $\tilde{\beta}_1 = 0.306$ | 0 | GEE | 0.301 | −0.005 | 0.103 | 0.102 | 0.103 | 0.948 | 0.949 |
| | 30 | PO | 0.300 | −0.006 | 0.106 | 0.106 | 0.106 | 0.948 | 0.949 |
| | 30 | MI | 0.302 | −0.004 | 0.106 | 0.106 | 0.106 | 0.947 | 0.949 |
| | 70 | PO | 0.295 | −0.011 | 0.114 | 0.114 | 0.114 | 0.948 | 0.949 |
| | 70 | MI | 0.301 | −0.005 | 0.115 | 0.113 | 0.113 | 0.945 | 0.946 |
| $\tilde{\beta}_2 = 0.637$ | 0 | GEE | 0.622 | −0.015 | 0.097 | 0.097 | 0.097 | 0.946 | 0.946 |
| | 30 | PO | 0.618 | −0.019 | 0.100 | 0.100 | 0.101 | 0.943 | 0.943 |
| | 30 | MI | 0.622 | −0.015 | 0.101 | 0.100 | 0.100 | 0.942 | 0.943 |
| | 70 | PO | 0.608 | −0.029 | 0.108 | 0.108 | 0.108 | 0.939 | 0.941 |
| | 70 | MI | 0.623 | −0.014 | 0.109 | 0.107 | 0.107 | 0.942 | 0.944 |

(Coef., "true" value of the coefficient; % Cens., percent of individuals subject to censoring prior to 5 years of follow-up. For methods: GEE, standard generalized estimating equation approach applied to uncensored version of the data; PO, pseudo-observation approach; MI, multiple-imputation approach. For remaining column headings: $\hat{\beta}$, average coefficient estimate; Bias, average $\hat{\beta} - \tilde{\beta}$; ESD, empirical standard deviation of $\hat{\beta}$; SE Unstr., the average estimated robust standard error using an unstructured working correlation matrix; SE Toepl., the average estimated robust standard error using a Toeplitz working correlation matrix; CP Unstr., empirical coverage probability for true coefficient based on 95% confidence interval using robust standard error with an unstructured working correlation matrix; CP Toepl., empirical coverage probability for true coefficient based on 95% confidence interval using robust standard error with an Toeplitz working correlation matrix.)

A Gaussian copula approach[45] is used to generate correlated exponential gap times in this section and in Section 3. This approach first simulates mean zero multivariate normal random variables $\{Q_{i1}, Q_{i2}, \ldots, Q_{i500}\}$ with variance one and 0.8 correlation between $Q_{ij}$ and $Q_{ij'}$, for $j \neq j'$; 500 was chosen to ensure that individuals would have at least 10 years of gap time history (5-year burn-in period, followed by 5 years of potential follow-up, subject to the previously described censoring mechanism). We then transform the multivariate normal random variables to multivariate uniform$(0, 1)$ and then multivariate exponential random variables via repeated applications of the inverse transform theorem. The exponential random variables that result from this process become the correlated gap times $\{G_{i1}, G_{i2}, \ldots, G_{i500}\}$, which are then converted into times-to-first event as described in Section 2.

Figure S5 in the Supplementary Materials shows the empirical average of $\log[\min\{2, T_i(t)\}]$ for each value of $Z_i = 0, 1, 2$ based on $N = 10\,000$ individuals with correlated exponential gap time histories as generated using the copula approach described above. As expected based on the stabilization of the survival curves seen in Figure S4 in the Supplementary Materials, the empirical average of $\log[\min\{\tau, T_i(t)\}]$ seems to stabilize successfully after the 5-year burn-in period. Results from $t = 5$ to $t = 8$ in this figure are averaged to provide nonparametric large sample estimates, $\tilde{\beta}$, of parameters in model (3)

$$E(\log[\min\{\tau, \mathcal{T}\}]|Z) = -0.677 + 0.306I(Z = 1) + 0.637I(Z = 2).$$

Finite sample properties of the uncensored case (where GEE is used) and censored cases (where PO and MI methods are used) shown in Table 2 are based on $N = 500$ individuals simulated to have an equal chance of following the simulated time-to-first-event longitudinal data structure governed by covariate values, $Z = 0, 1$, or 2. Results are laid out in a similar manner to that seen in Table 1, except that bias is defined in relation to the nonparametric large sample estimate, $\tilde{\beta}$ rather than a true $\beta$, since a closed-form value for the true $\beta$ is unavailable. Results are very comforting. The MI method slightly outperforms the PO method with slightly lower SEs and lower bias; bias results for the MI method are comparable to that seen in the uncensored case. CPs are close to the desired 95%. Parameter estimates corresponding to groups with shorter suggested simulation burn-in periods ($\hat{\beta}_0$ and $\hat{\beta}_1$) have slightly improved bias results compared to $\hat{\beta}_2$. Since we used the same 5-year burn-in period for all covariate values, the values of $\beta_0$ and $\beta_1$ had stabilized earlier in the burn-in period, as seen in Figure S5, and corresponding estimates may have benefited slightly from this feature of the simulation.

In the Supplementary Materials, for comparison, we also provide results if data from only the first event time is used to fit model (3), ie, the data consists of $\{X_i(t=0), \delta_i(t=0)\}, i = 1, \ldots, 500$. Both PO and MI methods in this special case follow those described by Liu et al.[42] Standard linear regression is used to fit model (3) in the setting with no censored data. Tables S2 and S3 summarize results corresponding to Tables 1 and 2, respectively, when only the first event time is used in the analysis. Table S4 shows the asymptotic relative efficiencies comparing our analyses versus the analyses with only the time-to-first-event, which vary from 1.179 to 1.880. As expected, there is a loss of efficiency when ignoring information after the first event time. This efficiency loss is most costly in the setting with independent gap times between events, but remains substantial when gap times are correlated.

## 6 | EXAMPLE

In this section, we use the proposed methods to analyze results from the azithromycin in COPD Trial. This study followed 1112 patients with a history of AEs for recurrent AEs after randomization to either placebo or 250 mg azithromycin daily. This trial ended with favorable results for the azithromycin arm,[1] based on an analysis of the time to first AE using the logrank test. Multivariable Cox proportional hazard analysis modeling time-to-first-exacerbation confirmed azithromycin benefit after adjustment for forced expiratory volume in one second ($FEV_1$), age, gender, smoking status, and study sites.

In our analysis, we estimate parameters in model (3) for $\tau = 6$ months and a longitudinal data structure, $\mathcal{T}$, measuring times-to-first-recurrent-event in follow-up windows starting at times $t = 0, 2, 4, 6$ months. Our selections of $\tau = 6$ and $a = 2$ months are based on the 6-month historic mean time-to-exacerbation in this patient population and recommendations from the work of Xia and Murray[25] that approximately 90% of recurrent events should be captured when spacing windows apart by one-third of a historic mean.

We present results from a univariate analysis that evaluates azithromycin versus placebo, a forest plot analysis of treatment effect in subgroups of interest, and a multivariable analysis of treatment effect that adjusts for age, gender, $FEV_1$, smoking status, and study site. We tested for and found no statistically significant interactions between follow-up window start times and treatment, indicating relatively stable patterns of treatment effect over time ($p > 0.41$). The Toeplitz working correlation structure gave a slightly lower QIC value compared to the unstructured working correlation structure in our multivariable model and was used in all models of the azithromycin data. All confidence intervals and p-values are based on robust sandwich estimation of variability.

Forest plots of univariate treatment effects, overall and by subgroup, are shown in Figure 2 for the PO (left panel) and MI (right panel) methods. Tabulated versions of these results are located in Table S5 of the Supplementary Materials. Treatment effects are displayed on the scale of $e^{\hat{\beta}}$ and can be interpreted as multiplicative increases (or decreases) on the time to first exacerbation over the next 6 months of follow-up. Overall, azithromycin is estimated to extend the time to the first exacerbation over a 6-month period by approximately 14% using either the PO or the MI method (95% CI approximately 5% to 24% longer, $p = 0.002$ for PO method and $p = 0.001$ for MI method). Stated as an absolute difference,

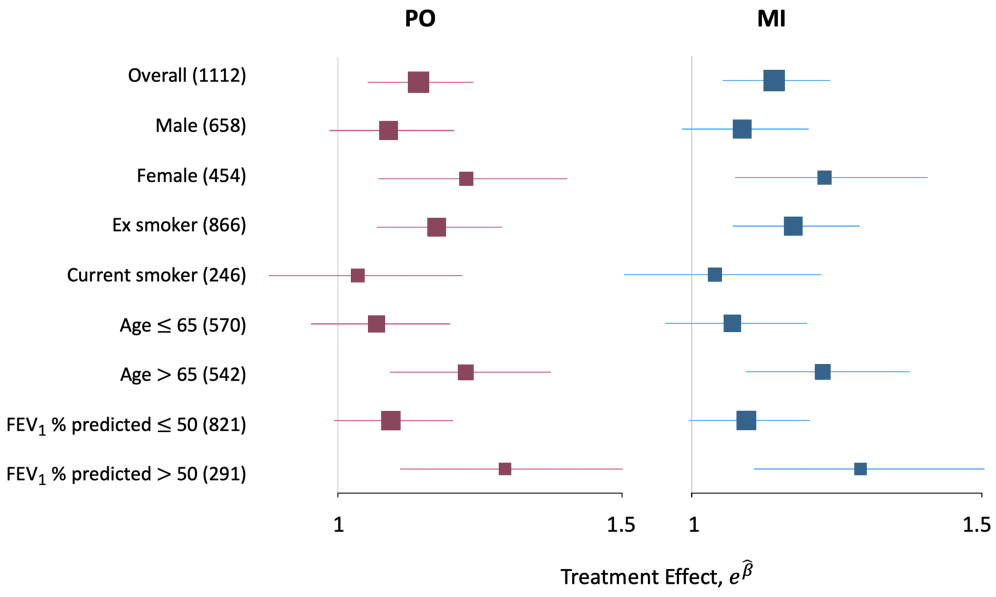


**FIGURE 2** Forest plot of univariate treatment effects, overall and by subgroups of interest [Colour figure can be viewed at wileyonlinelibrary.com]

**TABLE 3** Multivariable results using PO and MI methods. Displayed estimates are additionally adjusted for center (data not shown). (CI: confidence interval; PO: pseudo-observation; MI: multiple imputation)

| | PO | | | | MI | | | |
|---|---|---|---|---|---|---|---|---|
| | $e^{\hat{\beta}}$ | 95% CI | | P | $e^{\hat{\beta}}$ | 95% CI | | P |
| **Intercept** | 60.34 | 41.17 | 88.43 | <0.001 | 61.28 | 42.33 | 88.72 | <0.001 |
| **Azithromycin (vs placebo)** | 1.153 | 1.063 | 1.250 | 0.001 | 1.150 | 1.063 | 1.244 | <0.001 |
| **$FEV_1$ (per 10% predicted)** | 1.041 | 1.013 | 1.068 | 0.003 | 1.039 | 1.013 | 1.066 | 0.003 |
| **Age (per 10 years)** | 1.052 | 0.999 | 1.108 | 0.055 | 1.052 | 1.001 | 1.107 | 0.046 |
| **Male (vs female)** | 1.181 | 1.084 | 1.287 | <0.001 | 1.171 | 1.078 | 1.272 | <0.001 |
| **Current smoker (vs Ex)** | 1.074 | 0.967 | 1.193 | 0.184 | 1.071 | 0.967 | 1.186 | 0.188 |

there was an estimated 0.43 month increase in time-to-first-exacerbation for the azithromycin group compared to the placebo group over a 6-month period (ie, $e^{\hat{\beta}_0+\hat{\beta}_1} - e^{\hat{\beta}_0} \approx 0.43$ using either the PO or MI method).

Across the various subgroup analyses shown in Figure 2, the treatment benefit was most pronounced in COPD patients with better preserved lung function, that is, $FEV_1$ % of predicted > 50 % (approximately 29% longer time to first exacerbation in the next 6 months, 95% CI 10%-51% longer using PO method and 11%-50% longer using MI method). In general, point estimates shown for the PO and MI methods in Table S5 are very close to one another and 95% CI results for the methods are also close, but with slightly narrower CI widths using the MI approach. P-values for the MI method are also slightly smaller using the MI versus the PO method.

As seen in Table 3, the azithromycin group maintains its estimated treatment benefit when adjusted for confounders in a multivariable model using either the PO or MI method (approximately 15% longer time to first exacerbation in the next 6 months, 95% CI 6%-25% longer, $p = 0.001$, using PO method, and 6%-24% longer, $p < 0.001$, using MI method. Interactions between treatment and $FEV_1$, age, gender, smoking status other than study sites were tested and no significant interactions were found.

For comparison, Table S6 in the Supplementary Materials gives results from a multivariable proportional means/rates model,[15] which is an alternative attractive method for analysis of recurrent events data. Parameter estimates, $e^{\hat{\beta}}$, in Table S6 are interpreted in relation to the estimated number of exacerbations in the clinical trial. For instance, the azithromycin parameter estimate (0.815) indicates that azithromycin patients experienced only 81.5% of the exacerbations that placebo patients experienced after adjustment for confounders (95% CI 71.1%-93.4%, $p = 0.003$). If time-dependent covariates are used with the proportional means model, the interpretation of parameter estimates applies to the intensity rate of the recurrent events, rather than the mean counts. Scientifically, conclusions using either our method or the proportional means/rates model are consistent with one another, although parameter estimates have different interpretations and are not directly comparable. Statistical significance of the various parameters in their respective models are reasonably similar in this case, with slightly stronger significance for the treatment effect using our proposed model.

## 7 | DISCUSSION

In this manuscript, we take a fresh look at the manner in which recurrent event data is analyzed. By first restructuring the data into a censored longitudinal form and then transforming the data via PO or MI models into a complete data format, we are able to take advantage of existing software from longitudinal data analysis literature. Our model estimates time free from recurrence over a $\tau$-length follow-up period. In our opinion, this model gives a clear manner of assessing clinical and statistical significance of associations simultaneously. As with most longitudinal data, our method allows for either time-independent or dependent predictors. Compared to time-to-first-event analysis estimates of parameters from model (3), our proposed methods for estimating these parameters, using additional recurrent event information, are much more efficient.

We are unaware of anyone who has suggested modeling a censored longitudinal data structure as we propose for the recurrent events setting, along with estimation methods that are valid regardless of the correlation structure between recurrent events. We are equally unaware of any literature that has described how to effectively simulate correlated times-to-first-event that follow a stable distribution over time. Our simulation results confirm the success of the simulation procedure as well as the data analysis methodology.

Compared to time-to-first-event analyses, our methodology for the analysis of recurrent event data gains a great deal of efficiency with limited additional assumptions required. In the special case where treatment group is the only predictor, the test of the treatment group parameter is comparable to the two-sample test of Tayob and Murray.[46] In that paper,

power of the Tayob and Murray statistic for detecting treatment differences was compared to power using the proportional means/rates model with treatment group as the only predictor and compared favorably, particularly when gap times were correlated. More generally, parameter estimates from our model and the proportional means/rates model do not share a similar scale or interpretation and the linear predictor from these respective models reflects very different assumptions regarding the behavior of the recurrent events over time.

A novel contribution of our research is a better understanding of how times-to-first-event fit within the context of recurrent events data. For many chronic diseases, clinic visits are scheduled at regular intervals and a good understanding of short-term patient prognosis is desired at these visits. Since these visits are typically scheduled months in advance, they are more likely to coincide with a period of clinical stability than an emergent event. Summary statistics from a time-to-next event analysis seem more relevant than a gap-time analysis in this setting. Similarly, in clinical trials of chronic diseases where patients are stable at study entry, the first event experienced during the trial is not a true gap time unless event times are independent and memoryless; a feature of the data we outline clearly for the first time. Our method is especially relevant in these settings.

Alternatively, when patients are in the midst of a recurrent event episode and seeking guidance on their prognosis, a gap time analysis is more immediately relevant and appropriate. Gap time analyses are also very appropriate for clinical trials that enroll patients during emergent events.

A possible modification to our MI procedure would be to incorporate an additional bootstrap step, which experts in MI theory view as a draw from the parameter space, going as far as to say that imputation must include this step to be proper. For instance, a bootstrap sample could be selected with replacement from the original dataset and used to fit model (4), providing parameter estimate $\beta^{PO}$. For individual $i$ in the original dataset requiring a sup impute, a risk set $\mathcal{R}_i$ would then be taken from the bootstrapped sample who satisfy the two constraints defined in Section 4.2. Other steps of our procedure would proceed without further alteration. The argument for an additional bootstrap step of this nature is bolstered by noticeably improved CPs when this step is included in some cases.[40] In our own work with inverse probability transform imputation methods, we have not observed a sufficient improvement in CPs to justify the extra computing time needed to perform this extra step. In particular, CPs for the MI method seen in Tables 1 and 2 are satisfactorily close to the uncensored case CPs to justify skipping this step. Although purists will likely agree to disagree, we feel comfortable recommending our imputation algorithm, as is, given the CP results seen in simulation.

Compared to the MI approach, the PO approach is very easy to program and use in exploratory multivariable regression analysis. Because POs are estimated in a completely nonparametric manner, they may be estimated once and then used as if they were complete data throughout the remainder of model exploration. MI approaches that incorporate information from modeled covariates often require a new set of multiply imputed datasets to be created as additional important predictors or confounders are identified throughout the data analysis. The advantage of additional efficiency from the MI approach may be offset by the convenience and ease of the PO approach. One strategy would be to perform model exploration using the PO approach and then use the MI approach at the final stage of the analysis. However, the PO approach seems more than adequate as a standalone procedure with negligible loss of efficiency compared to the MI method and very quick results for the busy practitioner.

Further discussion of the choice of the user-defined parameter, $a$, in the special case where two-sample tests are of interest is given by Xia and Murray.[25] Our regression methodology allows for follow-up windows to start as frequently as every day ($a = 1$ day), in which case all recurrent events would be incorporated into the analysis. Xia and Murray,[25] however, recommend using $a =$ one-third of the historic mean time-to-recurrent event, capturing approximately 90% of the recurrent events in at least one follow-up window. Their recommendation is based on minimal loss of power to detect treatment effects and faster computing time for this larger value of $a$, which can be important when evaluating many models in our regression setting.

## ACKNOWLEDGEMENTS

The authors would like to thank the COPD Clinical Research Network, particularly our University of Michigan collaborator, Meilan Han, for the use of their data. We additionally thank Jeremy Taylor for useful feedback on our first manuscript draft.

## CONFLICT OF INTEREST

None declared.

## DATA AVAILABILITY STATEMENT

The R codes and data that support the findings of this study are openly available in GitHub at https://github.com/summerx0821/Regression-Recurrent-Event-Free-Time.[47]

## ORCID

*Meng Xia* https://orcid.org/0000-0002-9711-6215
*Nabihah Tayob* https://orcid.org/0000-0001-6088-167X

## REFERENCES

1. Albert RK, Connett J, Bailey WC, et al. Azithromycin for prevention of exacerbations of COPD. *N Engl J Med*. 2011;365(8):689-698.
2. Schwartz GG, Steg PG, Szarek M, et al. Alirocumab and cardiovascular outcomes after acute coronary syndrome. *N Engl J Med*. 2018;379(22):2097-2107. https://doi.org/10.1056/NEJMoa1801174
3. Wilcox MH, Gerding DN, Poxton IR, et al. Bezlotoxumab for prevention of recurrent clostridium difficile infection. *N Engl J Med*. 2017;376(4):305-317. https://doi.org/10.1056/NEJMoa1602615
4. DeKosky ST, Ikonomovic MD, Gandy S. Traumatic brain injury football, warfare, and long-term effects. *N Engl J Med*. 2010;363(14):1293-1296.
5. Frome EL, Kutner MH, Beauchamp JJ. Regression analysis of poisson-distributed data. *J Am Stat Assoc*. 1973;68(344):935-940.
6. Lawless J. Negative binomial and mixed Poisson regression. *Can J Stat*. 1987;15(3):209-225.
7. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*. 1992;34(1):1-14.
8. Greene WH. Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working paper. New York, NY: New York University; 1994.
9. Ozga A-K, Kieser M, Rauch G. A systematic comparison of recurrent event models for application to composite endpoints. *BMC Med Res Methodol*. 2018;18(1):2.
10. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat*. 1982:1100-1120.
11. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981;68(2):373-379.
12. Wei L-J, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc*. 1989;84(408):1065-1073.
13. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J Am Stat Assoc*. 1993;88(423):811-820.
14. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics*. 1995;37(2):158-168.
15. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J Royal Stat Soc B*. 2000;62(4):711-730.
16. Aalen OO, Husebye E. Statistical analysis of repeated events forming renewal processes. *Statist Med*. 1991;10(8):1227-1240.
17. Hougaard P. Frailty models for survival data. *Lifetime Data Anal*. 1995;1(3):255-273.
18. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics*. 2004;60(3):747-756.
19. Chang S-H. Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Anal*. 2004;10(2):175-190.
20. Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*. 2007;8(4):708-721.
21. Mazroui Y, Mathoulin-Pélissier S, MacGrogan G, Brouste V, Rondeau V. Multivariate frailty models for two types of recurrent events with a dependent terminal event: application to breast cancer data. *Biometrical Journal*. 2013;55(6):866-884.
22. Rogers JK, Yaroshinsky A, Pocock SJ, Stokar D, Pogoda J. Analysis of recurrent events with an associated informative dropout time: application of the joint frailty model. *Statist Med*. 2016;35(13):2195-2205.
23. Ding J, Sun L. Additive mixed effect model for recurrent gap time data. *Lifetime Data Anal*. 2017;23(2):223-253.
24. Cook RJ, Lawless J. *The Statistical Analysis of Recurrent Events*. New York, NY: Springer Science & Business Media; 2007.
25. Xia M, Murray S. Commentary on Tayob and Murray (2014) with a useful update pertaining to study design. *Biostatistics*. 2018;20(3):542-545.
26. Lin D, Sun W, Ying Z. Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika*. 1999;86(1):59-70.
27. Zeger SL, Liang K-Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986:121-130.
28. Westgate PM, Braun TM. The effect of cluster size imbalance and covariates on the estimation performance of quadratic inference functions. *Statist Med*. 2012;31(20):2209-2222.
29. Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Anal*. 2004;10(4):335-350.
30. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 2005;61(1):223-229.

31. Andersen PK, Klein JP. Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scand J Stat*. 2007;34(1):3-16.
32. Andrei A-C, Murray S. Regression models for the mean of the quality-of-life-adjusted restricted survival time using pseudo-observations. *Biometrics*. 2007;63(2):398-404.
33. Graw F, Gerds TA, Schumacher M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal*. 2009;15(2):241-255.
34. Xiang F, Murray S. Restricted mean models for transplant benefit and urgency. *Statist Med*. 2012;31(6):561-576.
35. Tayob N, Murray S. Statistical consequences of a successful lung allocation system – recovering information and reducing bias in models for urgency. *Statist Med*. 2017;36:2435-2451.
36. Quenouille MH. Approximate tests of correlation in time-series 3. *Math Proc Camb Philos Soc*. 1949;45:483-484.
37. Quenouille MH. Notes on bias in estimation. *Biometrika*. 1956;43(3/4):353-360.
38. Tukey J. Bias and confidence in not quite large samples. *Ann Math Stat*. 1958;29:614.
39. Faucett CL, Schenker N, Taylor JMG. Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics*. 2002;58(1):37-47.
40. Taylor JMG, Murray S, Hsu C-H. Survival estimation and testing via multiple imputation. *Stat Probab Lett*. 2002;58(3):221-232.
41. Hsu C-H, Taylor JMG, Murray S, Commenges D. Survival analysis using auxiliary variables via non-parametric multiple imputation. *Statist Med*. 2006;25(20):3503-3517.
42. Liu LX, Murray S, Tsodikov A. Multiple imputation based on restricted mean model for censored data. *Statist Med*. 2011;30(12):1339-1350.
43. Xiang F, Murray S, Liu X. Analysis of transplant urgency and benefit via multiple imputation. *Statist Med*. 2014;33(26):4655-4670.
44. Little RJ, Rubin DB. *Statistical Analysis With Missing Data*. New York, NY: John Wiley & Sons, Inc; 1986.
45. Li DX. On default correlation: a copula function approach. 1999. https://ssrn.com/abstract=187289
46. Tayob N, Murray S. Nonparametric tests of treatment effect based on combined endpoints for mortality and recurrent events. *Biostatistics*. 2014;16(1):73-83.
47. Xia M. Regression-Recurrent-Event-Free-Time. GitHub. 2019.


## SUPPORTING INFORMATION

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

**How to cite this article:** Xia M, Murray S, Tayob N. Regression analysis of recurrent-event-free time from multiple follow-up windows. *Statistics in Medicine*. 2020;39:1–15. https://doi.org/10.1002/sim.8385