Supplementary Materials for Regression Analysis of Recurrent-Event-Free Time from Multiple Follow-up Windows

Meng Xia[a], Susan Murray[a], Nabihah Tayob[b]

[a] *University of Michigan, Department of Biostatistics, Ann Arbor, MI 48109, USA*

[b] *Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA 02215, USA*

1 Appendix A: The Derivation of the Marginal Distribution of $T_i(t)$ With Independent Recurrent Event Times

In this appendix section we describe that for an individual i, when $G_{ij}, j = 1, \ldots, J_i$ are independently and identically distributed as exponential with intensity λ_i the marginal distribution of $T_i(t)$ for a fixed t is also distributed as exponential with hazard λ_i. This result is trivial for the case where $t = 0$, since $T_i(t = 0) \equiv G_{i1} \sim Exp(\lambda_i)$. For the case when $t > 0$,

$$
\begin{aligned}
&Pr\{T_i(t) > u\}\\
=&Pr\{T_{i1} > t + u\} + Pr\{T_{i1} \le t, T_{i2} > t + u\} + \cdots + Pr\{T_{iJ_i-1} \le t, T_{iJ_i} > t + u\}\\
=&Pr\{T_{i1} > t + u\} + \sum_{j=2}^{J_i} Pr\{T_{ij-1} \le t, T_{ij} > t + u\}\\
=&Pr\{G_{i1} > t + u\} + \sum_{j=2}^{J_i} Pr\{T_{ij-1} \le t, T_{ij-1} + G_{ij} > t + u\}\\
=&\int_{t+u}^{\infty} \lambda_i e^{-\lambda_i y} dy + \sum_{j=2}^{J_i} \int_0^t \int_{t+u-p}^{\infty} f(T_{ij-1} = p, G_{ij} = q) dq dp
\end{aligned}
$$

where $T_{ij-1} \perp G_{ij}$ by assumption. We know that $T_{ij-1} = \sum_{k=1}^{k=j-1} G_{ik}$. When $G_{ik} \overset{iid}{\sim} Exp(\lambda_i)$, $T_{ij-1} \sim Gamma(j-1, \lambda_i)$, where $j - 1$ is the shape parameter and λ_i is the rate parameter. Then,

$$
\begin{aligned}
&Pr\{T_i(t) > u\}\\
=&e^{-\lambda_i(t+u)} + \sum_{j=2}^{J_i} \int_0^t \int_{t+u-p}^{\infty} \text{pdf}_{Gamma(j-1,\lambda_i)}(T_{ij-1} = p)\text{pdf}_{Exp(\lambda_i)}(G_{ij} = q) dp dq
\end{aligned}
$$

where

$$
\text{pdf}_{Gamma(j-1,\lambda_i)}(T_{ij-1} = p) = \frac{\lambda_i^{j-1}}{\Gamma(j-1)} p^{j-2} e^{-\lambda_i p},
$$

$$\text{pdf}_{Exp(\lambda_i)}(G_{ij} = q) = \lambda_i e^{-\lambda_i q}.$$

So,

$$\begin{aligned}
&Pr\{T_i(t) > u\} \\
=&e^{-\lambda_i(t+u)} + \sum_{j=2}^{J_i} \int_0^t \frac{\lambda_i^{j-1}}{\Gamma(j-1)} p^{j-2} e^{-\lambda_i p} \int_{t+u-p}^{\infty} \lambda_i e^{-\lambda_i q} dp dq \\
=&e^{-\lambda_i(t+u)} + \sum_{j=2}^{J_i} \int_0^t \frac{\lambda_i^{j-1}}{\Gamma(j-1)} p^{j-2} e^{-\lambda_i(t+u)} dp \\
=&e^{-\lambda_i(t+u)} + \sum_{j=2}^{J_i} \frac{(\lambda_i t)^{j-1}}{\Gamma(j)} e^{-\lambda_i(t+u)} \\
=&\sum_{j=1}^{J_i} \frac{(\lambda_i t)^{j-1}}{\Gamma(j)} e^{-\lambda_i t} \times e^{-\lambda_i u}
\end{aligned}$$

When $J_i \to \infty$, by Taylor series,

$$\sum_{j-1=0}^{\infty} \frac{(\lambda_i t)^{j-1}}{(j-1)!} = e^{\lambda_i t}.$$

Therefore, $Pr\{T_i(t) > u\} = e^{-\lambda_i u}$, namely, $T_i(t) \sim Exp(\lambda_i)$.

2 Appendix B: Example Showing the Imputation of Event Times

Figure S1 shows the same example participant from the Azithromycin in COPD Trial shown in Figure 1 in the main manuscript. For this example participant, $\mathcal{S}_i = \{180, 240, 300\}$ so that the sup window starts at $t^{sup}(\mathcal{S}_i) = 300$. The sup impute becomes $\tilde{T}_i\{t^{sup}(\mathcal{S}_i)\} = \tilde{T}_i\{300\} = 65$. For the window starting at 240 days, the imputed time-to-first-event becomes $\tilde{T}_i\{240\} = \tilde{T}_i\{300\} + 300 - 240 = 125$, which is greater than the censored time-to-first-event that was observed for this window, $X_i(240) = 113$. Similarly for the window starting at 180 days, $\tilde{T}_i\{180\} = \tilde{T}_i\{300\} + 300 - 180 = 185$, which is greater than the censored time-to-first-event that was observed for this window, $X_i(180) = 173$.

3 Appendix C: Independent and Identically Distributed Weibull Gap Time Simulations, with Time-dependent Covariates

Our approach to recurrent event data analysis is to reformulate the recurrent event data into a longitudinal data format. Once the censored nature of the longitudinal data has been handled, via use of pseudo-observations or multiple imputation, traditional longitudinal model building occurs that includes exploration

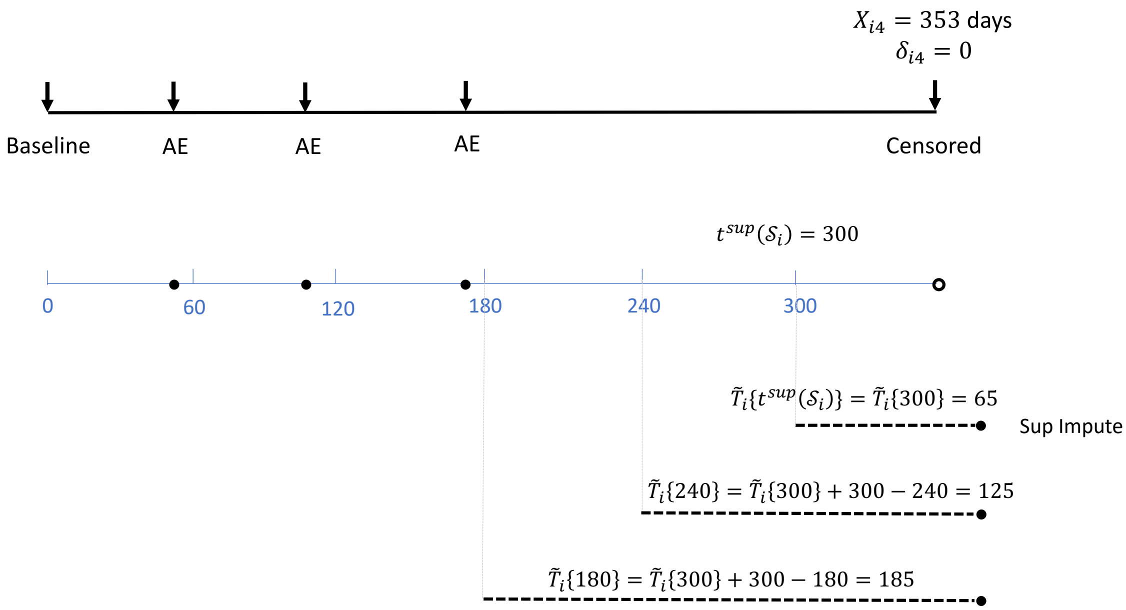

Figure S1: Example Showing the Imputation of Event Times

of time-dependent covariates in Model (3). As seen in Sections 3 and 5 of the main text, when (potentially correlated) gap times are exponentially distributed based on baseline covariates, Model (3) does not depend on t. For non-exponential underlying gap time distributions, Model (3) is likely to change with t, even when there are no measured covariates influencing the recurrent event behavior.

In this section, we present simulation results for fitting Model (3) in the scenario where gap times between events for individual, i, are independently and identically generated from a Weibull distribution with rate 1/3 and shape 2. For time-dependent covariates, $Z(t)$, we include indicator variables that reflect different window start times $\{t_2 = 1, \ldots, t_4 = 3\}$ with window start time $t_1 = 0$ as reference. 'True' model parameters for the case with $\tau = 2$ are nonparametrically estimated from a simulated dataset with $N = 10,000$ individuals, giving

$$E(log[min\{\tau, \mathcal{T}\}]) = 0.492 - 0.325I(t = 1) - 0.495I(t = 2) - 0.535I(t = 3).$$

Finite sample behavior shown in Table 3 is based on recurrent events from $N = 500$ individuals over 5 years of follow-up, with censoring distributions generated as described in Section 5 of the main text. All proposed approaches yield approximately unbiased estimates, coverage probabilities suitably close to 0.95 and estimated robust standard errors close to empirical standard deviations seen across simulations.

Figure S2 shows estimated recurrence-free time over the next 2 years and 95% confidence intervals for a typical simulated dataset, which shows that the 2-year restricted mean decreases over time.

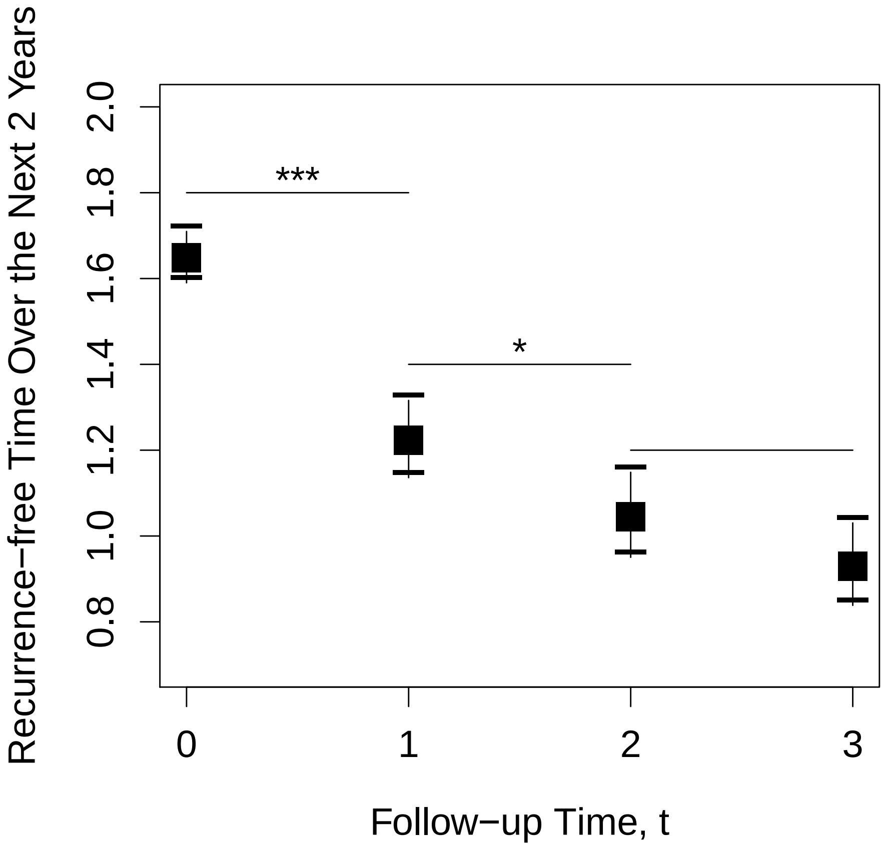

Figure S2: Estimated Recurrence-free Time Over the Next 2 Years and 95% Confidence Intervals for a Typical Simulated Dataset in the Scenario without Censoring.

***: p-value$<$ 0.001; *: p-value$<$ 0.05.

Coef.	**% Cens.**	**Method**	$\hat{\beta}$	**Bias**	**ESD**	**SE** Unstr.	**SE** Toepl.	**CP** Unstr.	**CP** Toepl.
	0	GEE	0.493	0.001	0.018	0.018	0.018	0.944	0.944
	30	PO	0.493	0.001	0.019	0.019	0.019	0.947	0.947
$\tilde{\beta}_0$=0.492	30	MI	0.493	0.001	0.019	0.019	0.019	0.944	0.944
	70	PO	0.493	0.001	0.020	0.020	0.020	0.941	0.941
	70	MI	0.493	0.001	0.020	0.020	0.020	0.937	0.937
	0	GEE	-0.326	-0.001	0.036	0.036	0.036	0.952	0.952
	30	PO	-0.327	-0.002	0.038	0.038	0.038	0.953	0.952
$\tilde{\beta}_1$=-0.325	30	MI	-0.326	-0.001	0.038	0.038	0.038	0.952	0.952
	70	PO	-0.327	-0.002	0.042	0.042	0.042	0.949	0.949
	70	MI	-0.325	0.000	0.042	0.042	0.043	0.950	0.951
	0	GEE	-0.494	0.001	0.048	0.048	0.048	0.946	0.946
	30	PO	-0.494	0.001	0.051	0.051	0.051	0.946	0.946
$\tilde{\beta}_2$=-0.495	30	MI	-0.493	0.002	0.051	0.051	0.051	0.943	0.942
	70	PO	-0.495	0.000	0.062	0.061	0.062	0.949	0.949
	70	MI	-0.492	0.003	0.062	0.060	0.060	0.942	0.942
	0	GEE	-0.533	0.002	0.048	0.048	0.048	0.948	0.948
	30	PO	-0.534	0.001	0.053	0.054	0.054	0.951	0.950
$\tilde{\beta}_3$=-0.535	30	MI	-0.533	0.002	0.053	0.053	0.053	0.949	0.948
	70	PO	-0.535	0.000	0.069	0.070	0.070	0.950	0.949
	70	MI	-0.534	0.001	0.069	0.069	0.069	0.946	0.946

Table S1: Simulated Finite Sample Performance for $N = 500$ Individuals with Time-dependent Covariates. Results are based on 10,000 iterates.

(Coef.: 'True' value of the coefficient; % Cens.: Percent of individuals subject to censoring prior to 5 years of follow-up;
For Methods, GEE: standard GEE 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.)

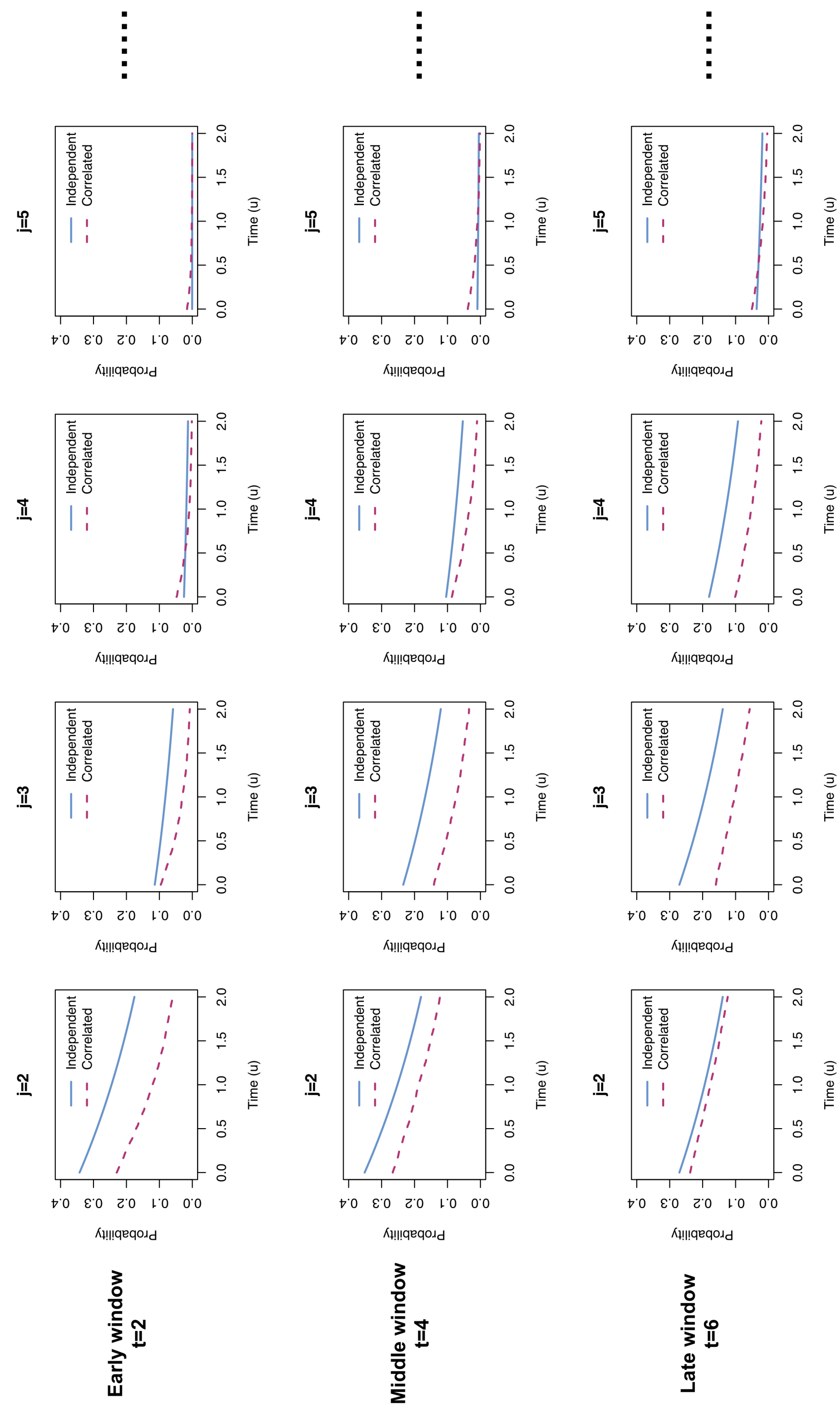

Figure S3: $Pr\{T_{ij-1} \leq t, T_{ij-1} + G_{ij} > t + u\}$ Over u for Specific t and j.

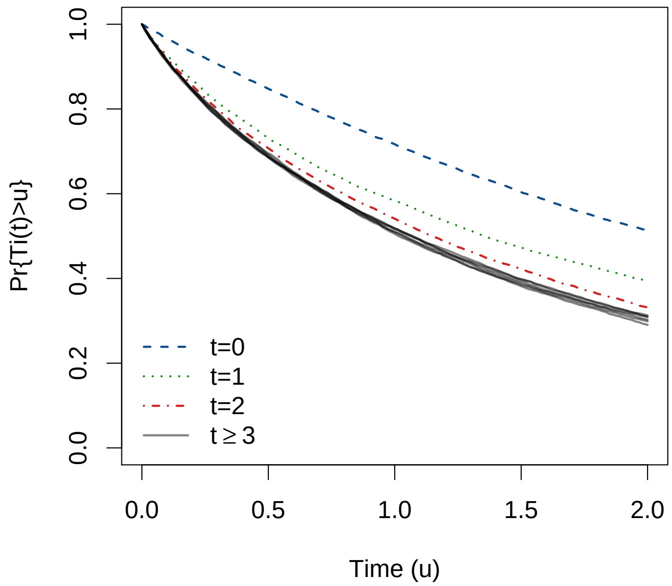

Figure S4: $Pr\{T_i(t) > u\}$ by Follow-up Window Start Times, t.

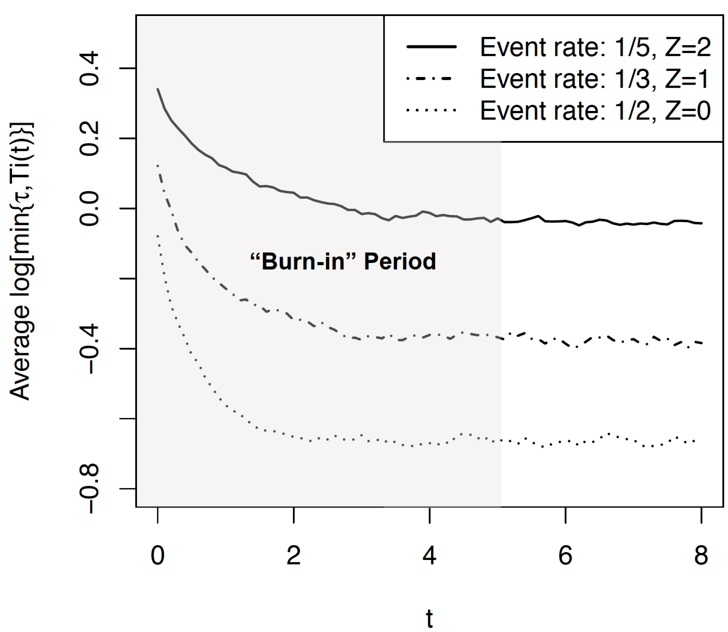

Figure S5: Empirical Average of $log[min\{2, T_i(t)\}]$. Based on N=10,000 individuals with correlated exponential gap time histories per curve; correlation is approximately 0.8; $t \in \{0, \ldots, 8\}$; a= 0.1 units apart. Curves seem to stabilize after the shaded burn-in period of 5 years.

Coef.	%Cens.	Method	$\hat{\beta}$	Bias	ESD	SE	CP
$\beta_0 = -0.7$	0	LM	-0.700	<0.001	0.108	0.102	0.934
	30	PO	-0.702	-0.002	0.109	0.103	0.935
	30	MI	-0.702	-0.002	0.109	0.103	0.935
	70	PO	-0.701	-0.001	0.113	0.105	0.930
	70	MI	-0.701	-0.001	0.113	0.105	0.930
$\beta_1 = 0.5$	0	LM	0.499	-0.001	0.100	0.100	0.947
	30	PO	0.500	<0.001	0.101	0.101	0.948
	30	MI	0.500	<0.001	0.101	0.101	0.948
	70	PO	0.500	<0.001	0.105	0.103	0.947
	70	MI	0.500	<0.001	0.105	0.103	0.947
$\beta_2 = 0.5$	0	LM	0.501	0.001	0.173	0.173	0.948
	30	PO	0.502	0.002	0.175	0.176	0.950
	30	MI	0.502	0.002	0.175	0.176	0.950
	70	PO	0.500	<0.001	0.180	0.179	0.950
	70	MI	0.500	<0.001	0.180	0.179	0.950

Table S2: Simulated Finite Sample Performance Using Only Time-to-first-event with Independent Times Between Recurrent Events.

(Coef.: True value of the coefficient; % Cens.: Percent of individuals subject to censoring prior to 5 years of follow-up;
For Methods, LM: standard linear model 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: the average estimated standard error; CP: empirical coverage probability for true coefficient based on 95% confidence interval.)

Coef.	%Cens.	Method	$\hat{\beta}$	Bias	ESD	SE	CP
$\tilde{\beta}_0$=-0.677	0	LM	-0.672	0.005	0.105	0.096	0.925
	30	PO	-0.674	0.003	0.106	0.097	0.926
	30	MI	-0.674	0.003	0.106	0.097	0.926
	70	PO	-0.673	0.004	0.105	0.099	0.937
	70	MI	-0.673	0.004	0.105	0.099	0.937
$\tilde{\beta}_1$=0.306	0	LM	0.306	<0.001	0.145	0.136	0.936
	30	PO	0.306	<0.001	0.146	0.138	0.935
	30	MI	0.306	<0.001	0.146	0.138	0.935
	70	PO	0.305	-0.001	0.144	0.140	0.943
	70	MI	0.305	-0.001	0.144	0.140	0.943
$\tilde{\beta}_2$=0.637	0	LM	0.632	-0.005	0.136	0.136	0.950
	30	PO	0.633	-0.004	0.138	0.138	0.949
	30	MI	0.633	-0.004	0.138	0.138	0.949
	70	PO	0.632	-0.005	0.139	0.140	0.951
	70	MI	0.632	-0.005	0.139	0.140	0.951

Table S3: Simulated Finite Sample Performance Using Only Time-to-first-event with Correlated Times Between Recurrent Events.

(Coef.: 'True' value of the coefficient; % Cens.: Percent of individuals subject to censoring prior to 5 years of follow-up;
For Methods, LM: standard linear model 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: the average estimated standard error; CP: empirical coverage probability for true coefficient based on 95% confidence interval.)

Coef.	%Cens.	Method	AREs (Recurrent vs. First Event) Indep Unstr.	Indep Toepl.	Corr Unstr.	Corr Toepl.
β_0	0	GEE/LM	1.789	1.789	1.280	1.280
	30	PO	1.689	1.689	1.244	1.244
	30	MI	1.717	1.717	1.260	1.260
	70	PO	1.500	1.500	1.193	1.179
	70	MI	1.544	1.522	1.207	1.193
β_1	0	GEE/LM	1.852	1.852	1.333	1.320
	30	PO	1.772	1.772	1.302	1.302
	30	MI	1.772	1.772	1.302	1.302
	70	PO	1.561	1.561	1.228	1.228
	70	MI	1.609	1.585	1.239	1.239
β_2	0	GEE/LM	1.880	1.880	1.402	1.402
	30	PO	1.796	1.778	1.380	1.366
	30	MI	1.814	1.796	1.380	1.380
	70	PO	1.584	1.584	1.296	1.296
	70	MI	1.613	1.613	1.308	1.308

Table S4: Asymptotic Relative Efficiencies (AREs) Comparing Analyses That Incorporate Recurrent Events Versus Only the Time-to-first-event.

(Coef.: Coefficient; % Cens.: Percent of individuals subject to censoring prior to 5 years of follow-up;
For Methods, GEE/LM: standard GEE approach applied to uncensored version of the data that reduces to the linear model (LM) in case of using only the time-to-first-event, PO: pseudo observation approach, MI: multiple imputation approach;
For remaining column headings, Indep: case with independently simulated gap times; Corr: case with correlated simulated gap times; Unstr.: unstructured working correlation matrix for the GEE method is used, robust variance is used for ARE calculation Toepl.: Toeplitz working correlation matrix for the GEE method is used, robust variance is used for ARE calculation.)

	PO				MI			
	$e^{\hat{\beta}}$	95% CI		P	$e^{\hat{\beta}}$	95% CI		P
Overall:	1.143	1.052	1.242	0.002	1.142	1.054	1.238	0.001
Sex:								
Male	1.088	0.983	1.204	0.104	1.087	0.984	1.201	0.100
Female	1.234	1.074	1.417	0.003	1.230	1.076	1.406	0.002
Smoking Status:								
Former	1.176	1.069	1.294	0.001	1.175	1.071	1.289	0.001
Current	1.032	0.875	1.218	0.707	1.040	0.884	1.223	0.639
Age:								
$\leq$ 65 years	1.067	0.949	1.198	0.278	1.069	0.955	1.198	0.246
$>$ 65 years	1.233	1.096	1.387	$<$0.001	1.226	1.093	1.374	$<$0.001
FEV1:								
$\leq$ 50 % predicted	1.095	0.994	1.208	0.067	1.094	0.995	1.202	0.063
$>$ 50 % predicted	1.291	1.104	1.510	0.001	1.289	1.107	1.502	0.001

Table S5: Subset Analyses Comparing Azithromycin versus Placebo Using Proposed PO and MI Approaches with a GEE Model. (CI: Confidence Interval; PO: Pseudo-observation; MI: Multiple Imputation.)

	Proportional Means/Rates Model			
	$e^{\hat{\beta}}$	95% CI		P
Azithromycin (vs. Placebo)	0.815	0.711	0.934	0.003
FEV_1 (per 10% Predicted)	0.918	0.875	0.962	<0.001
Age (per 10 Years)	0.927	0.853	1.007	0.072
Male (vs. Female)	0.814	0.709	0.935	0.004
Current Smoker (vs. Ex)	0.918	0.766	1.100	0.350

Table S6: Multivariable Results Using the Proportional Means/Rates Model Proposed by *Lin et al. (2000)*. Displayed estimates are additionally adjusted for center [data not shown]. (CI: confidence interval)