Supplemental Materials for "Random Forest for Dynamic Risk Prediction of Recurrent Events: A Pseudo-Observation Approach"

Abigail Loe[1,*], Susan Murray[1,+], and Zhenke Wu[1,+]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA; E-mail: *agloe@umich.edu; +: co-senior authors

A1 Kaplan-Meier Estimation

Using the notation from Section 2, let $t \in \mathcal{T}$, and define:

$$
\begin{aligned}
N_t(\tau) &= \sum_{i=1}^{n(t)} I[X_i(t) \le \tau, \delta_i(t) = 1] \\
dN_t(u) &= \sum_{i=1}^{n(t)} I[X_i(t) = u, \delta_i(t) = 1] \\
Y_t(u) &= \sum_{i=1}^{n(t)} I[X_i(t) \ge u]
\end{aligned}
$$

where $n(t)$ is the number of subjects available for follow-up at time point t. Briefly, $N_t(\tau)$ counts the number of events before horizon τ, $dN_t(u)$ counts the number of events at precisely u, and $Y_t(u)$ keeps track of the number of subjects at-risk for a recurrent event at time point u. Then the Kaplan-Meier estimator for probability of survival greater than τ at $t \in \mathcal{T}$ is defined as

$$
\widehat{P}[T_i(t) \ge \tau] = \prod_{i:X_i(t) \le \tau} \left\{1 - \frac{dN_t[X_i(t)]}{Y_t[X_i(t)]}\right\}.
$$

This probability can be estimated for each $t \in \mathcal{T}$ using the quantities calculated in Section 2.

A2 Model Calibration via Mean Squared Error

A calibration estimate of the m-XMT predictions averaged over all individuals, $i = 1, \dots, N$ and follow-up windows within individual, $t_0, \dots, t_{b_i}$ becomes

$$MSE = \frac{1}{\sum_{i=1}^{N}(b_i + 1)} \sum_{i=1}^{N} \sum_{j=0}^{b_i} [\widehat{S}_i^{\tau}(t_j) - \widehat{P}\{T_i(t_j) \geq \tau | \boldsymbol{Z}_i(t_j)\}]^2.$$

To obtain a similar calibration estimate for `RFRE.PO`, we replace the fitted values from the m-XMT model, $\widehat{P}\{T_i(t_j) \geq \tau | \boldsymbol{Z}_(t_j)\}$, with the predictions from the `RFRE.PO` method, $\widetilde{P}\{T_i(t_j) \geq \tau | \boldsymbol{Z}_i(t_j)\}$

A time-dependent MSE estimate pertaining to m-XMT predictions for the follow-up window starting at $t \in \mathcal{T}$ is:

$$MSE(t) = \frac{1}{n(t)} \sum_{i=1}^{n(t)} [\widehat{S}_i^{\tau}(t) - \widehat{P}\{T_i(t) \geq \tau | \boldsymbol{Z}_i(t)\}]^2.$$

Similarly, we can calculate $MSE(t)$ for `RFRE.PO` by replacing the predictions from an m-XMT model with the predictions from the `RFRE.PO` method, as described in Section 5.

A3 Additional Simulation Results: MSE

Figure S1 and Table S2 summarize MSE and $MSE(t), t \in \mathcal{T}$, simulation results for the `RFRE.PO` algorithm, the m-XMT (Main Effects) model, and the m-XMT (True) model. In general, the m-XMT (Main Effects) model has the highest MSE {Panel (A)}. In settings with correlation, the inclusion of history covariates decreases MSE for the m-XMT (Main Effects) model and the `RFRE.PO` algorithm. In all simulation settings, the `RFRE.PO` method outperforms the m-XMT (Main Effects) method, and upon the inclusion of history covariates, competes with the m-XMT (True) model in settings with highest correlation and lowest censoring. Examination of $MSE(t)$ reveals a similar stabilization as the stabilization observed in the C-statistic (Figure S1).

A4 Additional Simulation Results: Power and Type I Error

Tables S3 and S4 display the simulated proportion of times predictors were deemed statistically important in dynamic prediction methods using the `RFRE.PO` algorithm (based on the random forest permutation test) and the m-XMT (Main Effects) model (based on the Wald test), respectively. Columns Z_5 and Z_7 of these tables show low rates of deeming noise covariates as important predictors with more conservative rates seen using the `RFRE.PO` algorithm than the m-XMT (Main Effects) model. For the m-XMT (Main Effects) model, these rates reflect type I error rates for inclusion of these predictors in the model that are reasonably close to the nominal 0.05 type I level subject to simulation variability. For the `RFRE.PO` algorithm, these rates do not reflect the rate at which the predictors were used in the algorithm, but rather the rate at which these variables were significantly improving prediction outputs when included as an input in the algorithm.

Similarly, columns Z_1, Z_2, Z_3, Z_4, and Z_6 of these tables show the rate at which covariates associated with recurrent events in the m-XMT (True) model were deemed important in the `RFRE.PO` algorithm and the m-XMT (Main Effects) model in settings where (1) no history covariates were used, (2) $H_p^{(1)}(t)$ was included in the model or (3) $H_f^{(1)}(t)$ was used in the model. For the m-XMT (Main Effects) model, these rates correspond to power for detecting the association of these predictors in these 3 covariate history settings. For the `RFRE.PO` algorithm, these rates reflect the rates at which these variables were significantly improving prediction outputs when included as an input in the algorithm in these three history covariate settings.

When no history covariates are included as algorithm inputs, the `RFRE.PO` algorithm is more likely to deem these predictors as significantly helpful in making dynamic predictions when compared to the m-XMT (Main Effects) model. In settings where there is positive correlation between recurrent events and history covariates are included as inputs in the prediction algorithms, the relative importance of these history covariates tends to diminish rates of importance seen for Z_1, Z_2, Z_3, Z_4, Z_6 in Tables S3 and S4.

We were initially puzzled when looking at the importance of the covariate, t, in these tables, as t was not directly used in generating data via the m-XMT (True) model. Yet, the covariate, t, tends to show high importance in models with covariate history variables and correlation between recurrent event times. The explanation seems to lie with Panel B of Figure 3, where the C-statistic improves over the early follow-up windows when covariate history variables are used. This result seems to indicate an interaction between t and history variables in settings with correlated recurrent events that were not explicitly modeled in the data generation stage.

A5 Simulation Study: Additional Tables and Figures

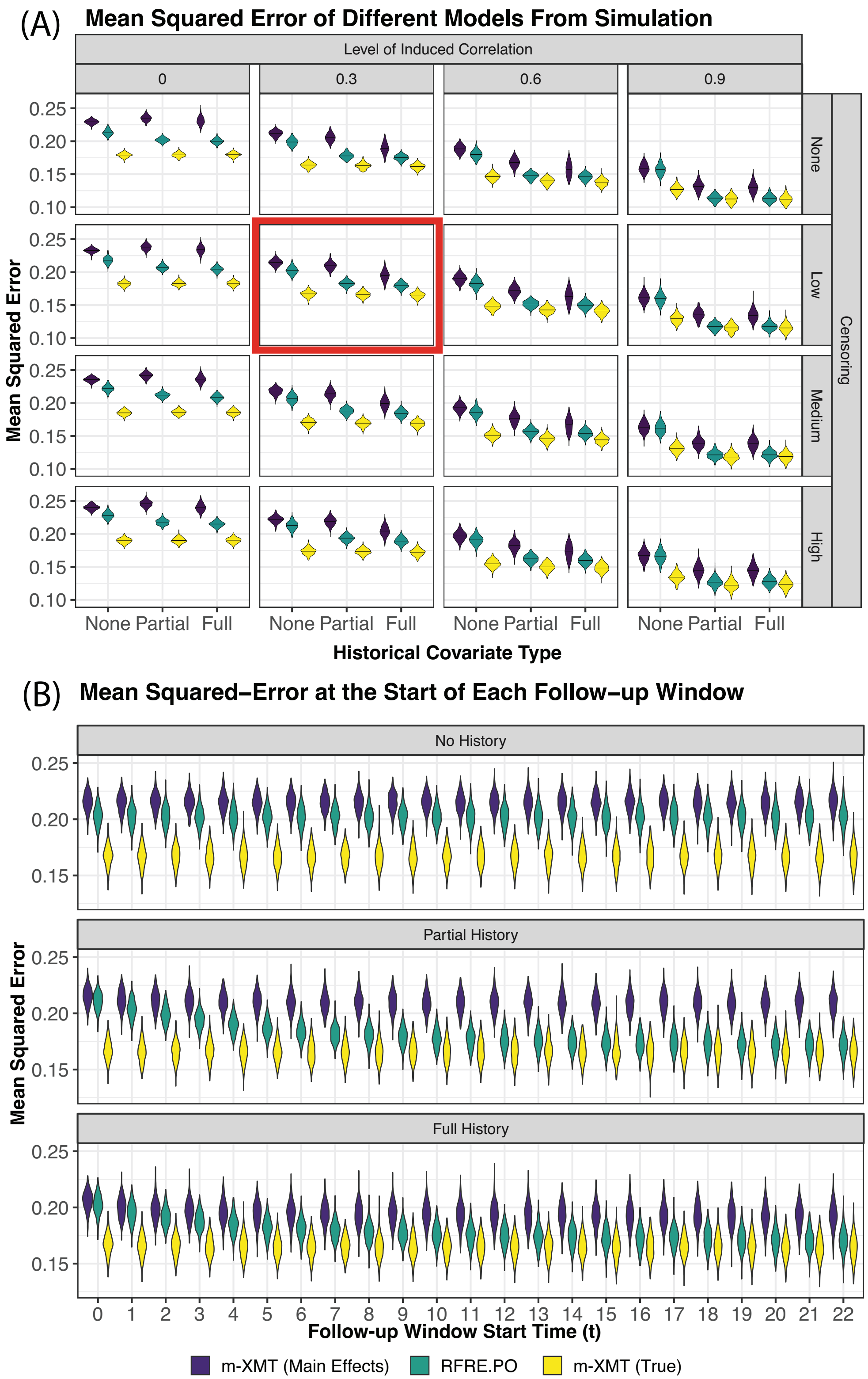

Figure S1: Panel (A) displays the mean squared error (MSE) for simulation settings displayed in Figure 3. Censoring increases from top to bottom, and the amount of induced correlation increases from left to right. Within a box, the quality of historical information increases from left to right. Panel (B) shows the MSE at the start of each follow-up window for the setting boxed in red in Panel (A). From top to bottom, more historical information is added in Panel (B), and the MSE corresponding to `RFRE.PO` in $H_p^{(1)}(t)$ settings. 500 replicates were performed, each with $n = 500$ subjects.

Table S1: Simulated C-statistics listed as Mean (ESD) for varying prediction algorithms when Z_1, Z_2 are correlated with each other, censoring percentages (none is 0%, light is 23%, moderate is 45%, and heavy is 63%), and gap-time correlations ($\rho = 0$, 0.3, 0.6, 0.9). 500 replicates were performed with $n = 500$ for each replicate, and the same data is reported in Figure 3.

	$\rho = 0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.9$
		No Censoring		
m-XMT (Main Effects)	0.534 (0.009)	0.533 (0.010)	0.533 (0.010)	0.531 (0.011)
`RFRE.PO`	0.605 (0.012)	0.596 (0.014)	0.584 (0.015)	0.563 (0.016)
m-XMT (True)	0.677 (0.006)	0.673 (0.007)	0.665 (0.008)	0.651 (0.010)
m-XMT (Main Effects) with $H_p^{(1)}(t)$	0.495 (0.021)	0.581 (0.018)	0.668 (0.012)	0.730 (0.008)
`RFRE.PO` with $H_p^{(1)}(t)$	0.633 (0.007)	0.653 (0.007)	0.685 (0.006)	0.712 (0.006)
m-XMT (True) with $H_p^{(1)}(t)$	0.677 (0.007)	0.674 (0.007)	0.685 (0.007)	0.711 (0.007)
m-XMT (Main Effects) with $H_f^{(1)}(t)$	0.522 (0.035)	0.648 (0.017)	0.687 (0.017)	0.732 (0.010)
`RFRE.PO` with $H_f^{(1)}(t)$	0.636 (0.008)	0.659 (0.008)	0.688 (0.008)	0.711 (0.007)
m-XMT (True) with $H_f^{(1)}(t)$	0.677 (0.006)	0.678 (0.007)	0.692 (0.010)	0.711 (0.010)
		Light Censoring		
m-XMT (Main Effects)	0.533 (0.009)	0.533 (0.009)	0.534 (0.010)	0.532 (0.011)
`RFRE.PO`	0.601 (0.015)	0.593 (0.014)	0.580 (0.017)	0.562 (0.017)
m-XMT (True)	0.677 (0.007)	0.672 (0.008)	0.664 (0.009)	0.650 (0.011)
m-XMT (Main Effects) with $H_p^{(1)}(t)$	0.496 (0.023)	0.574 (0.018)	0.668 (0.012)	0.724 (0.009)
`RFRE.PO` with $H_p^{(1)}(t)$	0.628 (0.008)	0.649 (0.007)	0.681 (0.007)	0.708 (0.006)
m-XMT (True) with $H_p^{(1)}(t)$	0.676 (0.006)	0.674 (0.007)	0.682 (0.008)	0.705 (0.008)
m-XMT (Main Effects) with $H_f^{(1)}(t)$	0.526 (0.032)	0.643 (0.017)	0.681 (0.019)	0.725 (0.011)
`RFRE.PO` with $H_f^{(1)}(t)$	0.634 (0.008)	0.656 (0.008)	0.685 (0.008)	0.708 (0.008)
m-XMT (True) with $H_f^{(1)}(t)$	0.677 (0.007)	0.676 (0.008)	0.689 (0.011)	0.706 (0.011)
		Moderate Censoring		
m-XMT (Main Effects)	0.534 (0.010)	0.535 (0.010)	0.535 (0.011)	0.532 (0.012)
`RFRE.PO`	0.597 (0.015)	0.590 (0.015)	0.577 (0.016)	0.558 (0.017)
m-XMT (True)	0.677 (0.007)	0.672 (0.008)	0.664 (0.009)	0.649 (0.012)
m-XMT (Main Effects) with $H_p^{(1)}(t)$	0.495 (0.022)	0.564 (0.018)	0.661 (0.013)	0.716 (0.011)
`RFRE.PO` with $H_p^{(1)}(t)$	0.625 (0.009)	0.646 (0.009)	0.677 (0.008)	0.704 (0.007)
m-XMT (True) with $H_p^{(1)}(t)$	0.677 (0.007)	0.674 (0.008)	0.681 (0.008)	0.700 (0.008)
m-XMT (Main Effects) with $H_f^{(1)}(t)$	0.529 (0.030)	0.638 (0.017)	0.676 (0.022)	0.719 (0.013)
`RFRE.PO` with $H_f^{(1)}(t)$	0.632 (0.009)	0.654 (0.008)	0.682 (0.009)	0.705 (0.008)
m-XMT (True) with $H_f^{(1)}(t)$	0.676 (0.007)	0.676 (0.008)	0.688 (0.011)	0.700 (0.010)
		Heavy Censoring		
m-XMT (Main Effects)	0.535 (0.010)	0.535 (0.011)	0.535 (0.012)	0.532 (0.012)
`RFRE.PO`	0.592 (0.015)	0.585 (0.016)	0.569 (0.017)	0.551 (0.017)
m-XMT (True)	0.677 (0.008)	0.673 (0.008)	0.664 (0.010)	0.649 (0.012)
m-XMT (Main Effects) with $H_p^{(1)}(t)$	0.498 (0.022)	0.557 (0.018)	0.641 (0.015)	0.707 (0.014)
`RFRE.PO` with $H_p^{(1)}(t)$	0.621 (0.009)	0.642 (0.008)	0.672 (0.008)	0.701 (0.007)
m-XMT (True) with $H_p^{(1)}(t)$	0.677 (0.008)	0.673 (0.008)	0.678 (0.009)	0.695 (0.009)
m-XMT (Main Effects) with $H_f^{(1)}(t)$	0.535 (0.031)	0.635 (0.019)	0.669 (0.025)	0.709 (0.017)
`RFRE.PO` with $H_f^{(1)}(t)$	0.628 (0.010)	0.652 (0.009)	0.678 (0.010)	0.701 (0.009)
m-XMT (True) with $H_f^{(1)}(t)$	0.676 (0.008)	0.676 (0.008)	0.685 (0.011)	0.695 (0.012)

Table S2: Simulated MSE listed as Mean (ESD) for varying prediction algorithms considered in Section 7, with censoring percentages (None, Light, Moderate, and Heavy), and gap-time correlations (ρ = 0, 0.3, 0.6, 0.9). 500 replicates were performed with $n = 500$ for each replicate.

	$\rho = 0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.9$
		No Censoring		
m-XMT (Main Effects)	0.229 (0.003)	0.212 (0.005)	0.188 (0.006)	0.159 (0.008)
`RFRE.PO`	0.213 (0.005)	0.199 (0.005)	0.180 (0.006)	0.157 (0.008)
m-XMT (True)	0.179 (0.003)	0.164 (0.004)	0.146 (0.005)	0.127 (0.006)
m-XMT (Main Effects) with $H_p^{(1)}(t)$	0.235 (0.005)	0.206 (0.006)	0.168 (0.007)	0.131 (0.008)
`RFRE.PO` with $H_p^{(1)}(t)$	0.202 (0.003)	0.178 (0.005)	0.147 (0.005)	0.114 (0.005)
m-XMT (True) with $H_p^{(1)}(t)$	0.179 (0.004)	0.163 (0.005)	0.140 (0.005)	0.112 (0.006)
m-XMT (Main Effects) with $H_f^{(1)}(t)$	0.231 (0.007)	0.189 (0.009)	0.157 (0.011)	0.130 (0.009)
`RFRE.PO` with $H_f^{(1)}(t)$	0.200 (0.004)	0.175 (0.004)	0.146 (0.005)	0.113 (0.005)
m-XMT (True) with $H_f^{(1)}(t)$	0.180 (0.004)	0.162 (0.004)	0.138 (0.005)	0.112 (0.006)
		Light Censoring		
m-XMT (Main Effects)	0.233 (0.004)	0.215 (0.005)	0.190 (0.006)	0.161 (0.008)
`RFRE.PO`	0.218 (0.005)	0.203 (0.006)	0.183 (0.007)	0.160 (0.008)
m-XMT (True)	0.182 (0.004)	0.167 (0.005)	0.148 (0.006)	0.129 (0.007)
m-XMT (Main Effects) with $H_p^{(1)}(t)$	0.238 (0.005)	0.210 (0.006)	0.172 (0.007)	0.135 (0.008)
`RFRE.PO` with $H_p^{(1)}(t)$	0.207 (0.004)	0.183 (0.005)	0.152 (0.006)	0.118 (0.006)
m-XMT (True) with $H_p^{(1)}(t)$	0.182 (0.004)	0.166 (0.005)	0.143 (0.006)	0.115 (0.006)
m-XMT (Main Effects) with $H_f^{(1)}(t)$	0.234 (0.006)	0.195 (0.008)	0.163 (0.011)	0.134 (0.009)
`RFRE.PO` with $H_f^{(1)}(t)$	0.204 (0.004)	0.180 (0.005)	0.150 (0.006)	0.118 (0.006)
m-XMT (True) with $H_f^{(1)}(t)$	0.183 (0.004)	0.166 (0.005)	0.141 (0.006)	0.116 (0.007)
		Moderate Censoring		
m-XMT (Main Effects)	0.236 (0.004)	0.218 (0.005)	0.193 (0.007)	0.163 (0.008)
`RFRE.PO`	0.222 (0.005)	0.207 (0.006)	0.186 (0.007)	0.162 (0.008)
m-XMT (True)	0.185 (0.004)	0.170 (0.005)	0.151 (0.006)	0.131 (0.007)
m-XMT (Main Effects) with $H_p^{(1)}(t)$	0.242 (0.005)	0.214 (0.006)	0.172 (0.007)	0.139 (0.008)
`RFRE.PO` with $H_p^{(1)}(t)$	0.212 (0.004)	0.188 (0.005)	0.156 (0.006)	0.122 (0.006)
m-XMT (True) with $H_p^{(1)}(t)$	0.186 (0.004)	0.169 (0.005)	0.146 (0.006)	0.119 (0.006)
m-XMT (Main Effects) with $H_f^{(1)}(t)$	0.237 (0.006)	0.200 (0.008)	0.169 (0.012)	0.139 (0.009)
`RFRE.PO` with $H_f^{(1)}(t)$	0.209 (0.004)	0.184 (0.005)	0.154 (0.006)	0.122 (0.006)
m-XMT (True) with $H_f^{(1)}(t)$	0.185 (0.004)	0.168 (0.005)	0.144 (0.006)	0.119 (0.007)
		Heavy Censoring		
m-XMT (Main Effects)	0.240 (0.004)	0.221 (0.005)	0.197 (0.008)	0.167 (0.009)
`RFRE.PO`	0.228 (0.005)	0.212 (0.006)	0.191 (0.008)	0.166 (0.009)
m-XMT (True)	0.190 (0.004)	0.173 (0.005)	0.155 (0.006)	0.135 (0.007)
m-XMT (Main Effects) with $H_p^{(1)}(t)$	0.246 (0.005)	0.219 (0.006)	0.182 (0.008)	0.145 (0.009)
`RFRE.PO` with $H_p^{(1)}(t)$	0.218 (0.004)	0.194 (0.005)	0.162 (0.006)	0.127 (0.007)
m-XMT (True) with $H_p^{(1)}(t)$	0.190 (0.004)	0.174 (0.005)	0.151 (0.006)	0.123 (0.007)
m-XMT (Main Effects) with $H_f^{(1)}(t)$	0.240 (0.006)	0.205 (0.008)	0.173 (0.011)	0.145 (0.01)
`RFRE.PO` with $H_f^{(1)}(t)$	0.215 (0.004)	0.189 (0.006)	0.160 (0.007)	0.128 (0.007)
m-XMT (True) with $H_f^{(1)}(t)$	0.190 (0.004)	0.172 (0.006)	0.148 (0.007)	0.124 (0.008)

Table S3: Simulated proportion of times that the corresponding covariate appeared as statistically significant according to the Wald test for the m-XMT (Main Effects) method. History covariates considered were of the type τ^*-restricted mean survival time in previous windows, where τ^* is $\frac{1}{12}$ years (described in further detail in Section 2.1). In the $H_p^{(1)}(t)$ case, results were from $m = 10$ multiply imputed datasets combined via Rubin's rules. 500 replicates were performed with $n = 500$ for each replicate. Covariates with signal (Z_1, Z_2, Z_3, Z_4, Z_6) are displayed first, and noise covariates are shown second (Z_5, Z_7). Within a censoring and correlation designation, quality of historical information included in the model increases, from top to bottom.

		Signal					Noise		History Covariates Included		
Censoring	ρ	Z_1	Z_2	Z_3	Z_4	Z_6	Z_5	Z_7	t	$H_p^{(1)}(t)$	$H_f^{(1)}(t)$
None	0.0	0.642	0.370	0.078	0.056	0.380	0.054	0.066	0.050	–	–
None	0.0	0.641	0.293	0.050	0.077	0.378	0.054	0.085	0.046	0.402	–
None	0.0	0.646	0.372	0.078	0.060	0.380	0.060	0.066	0.046	–	0.346
None	0.3	0.624	0.230	0.082	0.044	0.388	0.060	0.056	0.064	–	–
None	0.3	0.635	0.232	0.076	0.057	0.374	0.052	0.052	0.071	0.431	–
None	0.3	0.528	0.222	0.074	0.050	0.248	0.054	0.040	0.358	–	0.998
None	0.6	0.662	0.200	0.062	0.056	0.460	0.070	0.072	0.044	–	–
None	0.6	0.495	0.168	0.060	0.071	0.304	0.043	0.092	0.522	1.000	–
None	0.6	0.470	0.172	0.062	0.054	0.190	0.092	0.072	0.672	–	1.000
None	0.9	0.592	0.134	0.098	0.082	0.474	0.066	0.066	0.066	–	–
None	0.9	0.381	0.165	0.072	0.036	0.137	0.065	0.065	0.986	1.000	–
None	0.9	0.336	0.122	0.070	0.072	0.148	0.066	0.050	0.772	–	1.000
Low	0.0	0.573	0.244	0.074	0.046	0.299	0.072	0.044	0.062	–	–
Low	0.0	0.598	0.274	0.074	0.095	0.385	0.054	0.051	0.071	0.439	–
Low	0.0	0.581	0.246	0.082	0.044	0.303	0.066	0.046	0.066	–	0.289
Low	0.3	0.592	0.206	0.072	0.060	0.398	0.054	0.040	0.052	–	–
Low	0.3	0.546	0.272	0.043	0.060	0.301	0.073	0.043	0.066	0.222	–
Low	0.3	0.504	0.192	0.068	0.062	0.238	0.060	0.036	0.230	–	0.998
Low	0.6	0.636	0.212	0.086	0.060	0.478	0.052	0.066	0.056	–	–
Low	0.6	0.487	0.170	0.057	0.073	0.243	0.053	0.097	0.413	1.000	–
Low	0.6	0.474	0.192	0.072	0.074	0.226	0.068	0.070	0.528	–	1.000
Low	0.9	0.582	0.168	0.096	0.094	0.500	0.054	0.064	0.052	–	–
Low	0.9	0.351	0.120	0.060	0.043	0.187	0.060	0.060	0.967	1.000	–
Low	0.9	0.338	0.150	0.070	0.076	0.176	0.064	0.078	0.582	–	1.000
Medium	0.0	0.618	0.250	0.068	0.060	0.358	0.094	0.088	0.080	–	–
Medium	0.0	0.580	0.279	0.061	0.065	0.370	0.059	0.057	0.055	0.475	–
Medium	0.0	0.622	0.248	0.066	0.064	0.370	0.094	0.084	0.084	–	0.236
Medium	0.3	0.624	0.250	0.072	0.074	0.372	0.066	0.060	0.066	–	–
Medium	0.3	0.599	0.246	0.068	0.054	0.393	0.086	0.052	0.072	0.132	–
Medium	0.3	0.508	0.230	0.080	0.080	0.234	0.070	0.064	0.172	–	0.994
Medium	0.6	0.620	0.160	0.078	0.074	0.464	0.060	0.058	0.060	–	–

Continued on next page

Table S3: (continued)

Censoring	ρ	Signal Z_1	Z_2	Z_3	Z_4	Z_6	Noise Z_5	Z_7	History Covariates Included t	$H_p^{(1)}(t)$	$H_f^{(1)}(t)$
Medium	0.6	0.485	0.190	0.080	0.052	0.251	0.064	0.052	0.289	1.000	–
Medium	0.6	0.442	0.166	0.074	0.070	0.236	0.072	0.054	0.410	–	1.000
Medium	0.9	0.582	0.182	0.076	0.062	0.488	0.042	0.062	0.052	–	–
Medium	0.9	0.412	0.151	0.050	0.072	0.207	0.062	0.054	0.922	1.000	–
Medium	0.9	0.376	0.176	0.068	0.064	0.190	0.042	0.064	0.466	–	1.000
High	0.0	0.580	0.254	0.060	0.066	0.322	0.048	0.054	0.058	–	–
High	0.0	0.584	0.263	0.066	0.078	0.400	0.056	0.068	0.066	0.436	–
High	0.0	0.584	0.246	0.062	0.064	0.322	0.048	0.058	0.056	–	0.244
High	0.3	0.562	0.260	0.080	0.078	0.358	0.064	0.076	0.084	–	–
High	0.3	0.583	0.228	0.070	0.060	0.315	0.040	0.114	0.068	0.044	–
High	0.3	0.462	0.228	0.078	0.078	0.238	0.058	0.078	0.146	–	0.988
High	0.6	0.570	0.198	0.074	0.076	0.424	0.088	0.048	0.068	–	–
High	0.6	0.536	0.216	0.074	0.064	0.304	0.056	0.058	0.198	0.984	–
High	0.6	0.428	0.174	0.082	0.074	0.208	0.064	0.044	0.302	–	1.000
High	0.9	0.570	0.160	0.088	0.076	0.460	0.064	0.066	0.068	–	–
High	0.9	0.433	0.122	0.064	0.068	0.230	0.066	0.034	0.770	1.000	–
High	0.9	0.390	0.154	0.076	0.082	0.226	0.058	0.070	0.276	–	1.000

Table S4: Simulated proportion of times that the corresponding covariate appeared as statistically significant according to the random forest permutation test. 500 replicates were performed, with $n = 500$ subjects in each replicate. History covariates considered were of the type τ^*-restricted mean survival time in previous windows, where τ^* is $\frac{1}{12}$ years (described in further detail in Section 2.1). In the $H_p^{(1)}(t)$ case, results were from $m = 10$ multiply imputed datasets, combined via Rubin's rules. Covariates with signal (Z_1, Z_2, Z_3, Z_4, and Z_6) are displayed first, and noise covariates are shown second (Z_5, Z_7). Within a censoring and correlation designation, quality of historical information included in the model increases, from top to bottom.

		Signal					Noise		History Covariates Included		
Censoring	ρ	Z_1	Z_2	Z_3	Z_4	Z_6	Z_5	Z_7	t	$H_p^{(1)}(t)$	$H_f^{(1)}(t)$
None	0.00	1.000	1.000	0.954	0.066	1.000	0.050	0.050	0.062	–	–
None	0.0	1.000	0.552	0.008	0.000	1.000	0.000	0.000	0.538	1.000	–
None	0.00	1.000	0.988	0.866	0.050	1.000	0.038	0.036	0.992	–	1.000
None	0.3	1.000	0.994	0.838	0.052	1.000	0.044	0.034	0.046	–	–
None	0.3	1.000	0.332	0.006	0.002	1.000	0.000	0.002	0.994	1.000	–
None	0.3	1.000	0.942	0.578	0.032	1.000	0.044	0.038	0.996	–	1.00
None	0.6	1.000	0.954	0.610	0.048	1.000	0.038	0.042	0.046	–	–
None	0.6	1.000	0.223	0.010	0.026	1.000	0.006	0.000	1.000	1.000	–
None	0.6	1.000	0.808	0.356	0.032	1.000	0.036	0.036	1.000	–	1.00
None	0.9	1.000	0.792	0.334	0.032	1.000	0.032	0.034	0.046	–	–
None	0.9	1.000	0.123	0.012	0.127	0.998	0.002	0.006	1.000	1.000	–
None	0.9	1.000	0.488	0.136	0.034	1.000	0.034	0.026	1.000	–	1.00
Low	0.0	1.000	1.000	0.890	0.044	1.000	0.020	0.030	0.022	–	–
Low	0.0	1.000	0.413	0.004	0.000	1.000	0.000	0.000	0.451	1.000	–
Low	0.0	1.000	0.970	0.739	0.038	1.000	0.014	0.026	0.824	–	1.00
Low	0.3	1.000	0.982	0.780	0.032	1.000	0.038	0.044	0.038	–	–
Low	0.3	1.000	0.280	0.008	0.002	1.000	0.000	0.000	0.976	1.000	–
Low	0.3	1.000	0.884	0.476	0.028	1.000	0.028	0.030	0.840	–	1.00
Low	0.6	1.000	0.938	0.506	0.018	1.000	0.016	0.040	0.042	–	–
Low	0.6	1.000	0.172	0.010	0.024	1.000	0.002	0.002	1.000	1.000	–
Low	0.6	1.000	0.712	0.260	0.022	1.000	0.016	0.040	0.948	–	1.00
Low	0.9	1.000	0.716	0.264	0.052	1.000	0.016	0.034	0.048	–	–
Low	0.9	0.996	0.118	0.006	0.104	1.000	0.008	0.024	1.000	1.000	–
Low	0.9	1.000	0.424	0.122	0.022	1.000	0.024	0.018	0.996	–	1.000
Medium	0.0	1.000	0.998	0.832	0.032	1.000	0.032	0.030	0.042	–	–
Medium	0.0	1.000	0.303	0.004	0.000	1.000	0.000	0.000	0.403	1.000	–
Medium	0.0	1.000	0.926	0.618	0.034	1.000	0.042	0.026	0.696	–	1.00
Medium	0.3	1.000	0.982	0.694	0.038	1.000	0.026	0.030	0.070	–	–
Medium	0.3	1.000	0.222	0.010	0.006	1.000	0.000	0.000	0.967	1.000	–
Medium	0.3	1.000	0.818	0.406	0.032	1.000	0.022	0.012	0.638	–	1.00
Medium	0.6	1.000	0.898	0.448	0.036	1.000	0.020	0.022	0.064	–	–

Continued on next page

Table S4: (continued)

Censoring	ρ	Signal					Noise		History Covariates Included		
		Z_1	Z_2	Z_3	Z_4	Z_6	Z_5	Z_7	t	$H_p^{(1)}(t)$	$H_f^{(1)}(t)$
Medium	0.6	1.000	0.161	0.010	0.024	1.000	0.000	0.000	1.000	1.000	–
Medium	0.6	1.000	0.622	0.196	0.020	1.000	0.012	0.020	0.824	–	1.00
Medium	0.9	1.000	0.688	0.258	0.028	1.000	0.016	0.016	0.064	–	–
Medium	0.9	1.000	0.113	0.014	0.138	1.000	0.006	0.023	1.000	1.000	–
Medium	0.9	1.000	0.440	0.112	0.034	1.000	0.020	0.018	0.926	–	1.00
High	0.0	1.000	0.966	0.688	0.036	1.000	0.024	0.020	0.060	–	–
High	0.0	1.000	0.211	0.000	0.000	1.000	0.000	0.000	0.310	1.000	–
High	0.0	1.000	0.836	0.490	0.038	1.000	0.022	0.022	0.582	–	1.00
High	0.3	1.000	0.946	0.506	0.028	1.000	0.014	0.016	0.098	–	–
High	0.3	1.000	0.192	0.004	0.002	1.000	0.000	0.000	0.953	1.000	–
High	0.3	1.000	0.706	0.268	0.018	1.000	0.012	0.016	0.590	–	1.000
High	0.6	1.000	0.812	0.368	0.014	1.000	0.022	0.026	0.094	–	–
High	0.6	1.000	0.126	0.010	0.024	1.000	0.002	0.006	1.000	1.000	–
High	0.6	1.000	0.524	0.138	0.020	1.000	0.018	0.026	0.716	–	1.000
High	0.9	1.000	0.602	0.226	0.016	1.000	0.016	0.014	0.086	–	–
High	0.9	0.998	0.106	0.033	0.110	0.994	0.004	0.018	1.000	1.000	–
High	0.9	1.000	0.314	0.094	0.018	1.000	0.022	0.012	0.818	–	1.000

A6 Tables and Figures for the Analysis of the Azithromycin for the Prevention of COPD Cohort

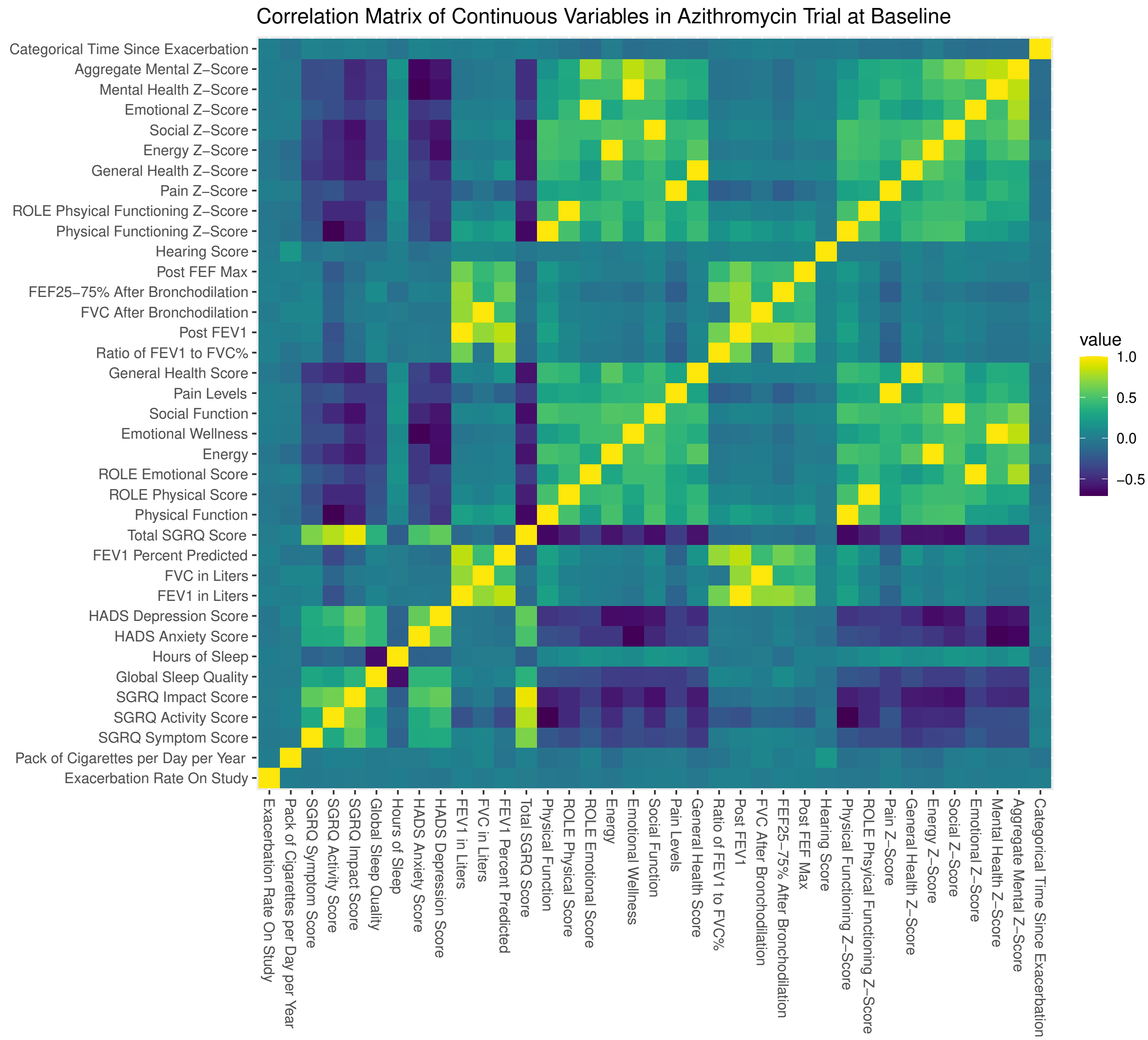

Figure S2: A correlation heat map of continuous and ordinal variables in the Azithromycin for the Prevention of COPD cohort, measured at baseline. Variables derived from each other have instances of particularly high positive or negative correlation. These correlation patterns make random forest a more attractive modeling alternative compared to semi-parametric and parametric methods.

Table S5: m-XMT parameter estimates, based on covariates selected by Xia et al. (2020), applied to censored longitudinal data with $\tau = 180$, $a = 30$.

Covariate	$e^{\widehat{\beta}\dagger}$	95% CI	p
Azithromycin (versus placebo)	1.43	(1.14, 1.79)	**<0.01**
Smoker at baseline (versus non-smoker)	1.07	(0.80, 1.43)	0.66
FEV_1 in liters at baseline	1.32	(1.05, 1.66)	**0.02**
Male gender (versus female)	1.36	(1.07, 1.74)	**0.01**
Age in decades	1.14	(0.99, 1.32)	0.07

† $e^{\widehat{\beta}}$ is the estimated odds ratio for being event-free during a 180-day follow-up period corresponding to a 1-unit increase in the (continuous) predictor, or the presence of a binary covariate, adjusted for other predictors.
FEV_1 is forced expiratory volume in one second, a clinical measure of lung function.

Table S6: Parameter estimates from the m-XMT (Forward Selection) model based on the first of $m = 10$ imputed datasets. For the analysis, we created three history variables, $H_f(t), H_p^{(2)}(t)$ and $H_p^{(3)}(t)$. $H_f(t)$ was an indicator if the subject had experienced a severe exacerbation on study time prior to time t. $H_p^{(2)}(t)$ was assumed to take a categorical form based on whether the participant had an exacerbation in the past zero to 31 days ($H_p^{(2)}(t) = 4$), 31-92 days ($H_p^{(2)}(t) = 3$), 93-182 days ($H_p^{(2)}(t) = 2$), 183 days to 365 days ($H_p^{(2)}(t) = 1$), or finally, if that time was more than 365 days or never ($H_p^{(2)}(t) = 0$). $H_p^{(3)}(t)$ is the estimated 30-day event rate using all follow-up prior to time t. A very recent exacerbation, bronchitis, gender, treatment, the ratio of FEV to FVC %predicted, SGRQ symptoms, oxygen use, recent hospitalization, rate of exacerbations, and use of ICS medication remain significant across the $m = 10$ imputed datasets.

Covariate	$e^{\widehat{\beta}}$ †	95% CI	p
Categorical $H_{p,i}^{(2)}(t) = 0$ (versus 1, 2, 3 or 4)	2.32	(2.01, 2.68)	<**0.01**
Hospitalization in the year before $t = 0$ (versus not)	0.70	(0.53, 0.92)	**0.01**
Corticosteroid use in the year before $t = 0$ (versus no use)	0.61	(0.40, 0.94)	**0.03**
$H_{f,i}(t) = 1$ (versus $H_{f,i}(t) = 0$)	1.81	(1.23, 2.68)	**0.01**
Time t (in months)	0.96	(0.93, 1.00)	**0.02**
Male gender (versus female gender)	1.69	(1.28, 2.23)	<**0.01**
Race = Black (versus non-Black)	0.57	(0.37, 0.89)	**0.01**
Sudden feelings of panic at time t (versus no such feelings)	0.84	(0.74, 0.95)	**0.01**
Sudden feelings of "butterfly" fright at time t (versus no such feelings)	0.89	(0.78, 1.01)	0.08
Seen a mental health provider at time t (versus has not seen a provider)	0.75	(0.61, 0.93)	**0.01**
Has pain at time t (versus no pain)	0.92	(0.86, 0.98)	**0.03**
St. George's Respiratory Questionnaire symptoms score at time t	0.99	(0.99, 1.00)	**0.01**
Antiarrhythmic use at time t (versus no use)	0.07	(0.00, 6.56)	0.25
FEV_1/FVC % predicted at time t	1.01	(1.00, 1.02)	<**0.01**
Azithromycin treatment group (versus placebo)	0.65	(0.49, 0.85)	<**0.01**
Inhaled corticosteroid or long-acting beta-agonist use at time t	1.28	(1.03, 1.60)	**0.02**
Other baseline medications (versus no miscellaneous medications)	0.15	(0.03, 0.81)	**0.02**
Supplemental oxygen use at time t (versus no supplemental oxygen)	0.55	(0.41, 0.73)	<**0.01**
Mixed beta-blocker use at time t (versus selective or no beta-blocker use)	3.79	(1.08, 13.24)	**0.04**
Bronchiectasis diagnosis at time t (versus no such diagnosis)	0.64	(0.48, 0.85)	<**0.01**

†: $e^{\widehat{\beta}}$ is the estimated odds ratio for being event-free during a 180-day follow-up period corresponding to a 1-unit increase in the (continuous) predictor, or the presence of a binary covariate, adjusted for other predictors.

FEV_1 is forced expiratory volume in one second and FVC is forced vital capacity, clinical measures of lung function.

SGRQ is the Saint George's Respiratory Questionnaire, a patient reported outcome measuring quality of life.

Table S7: P-values from the `RFRE.PO` permutation test applied to the out-of-bag-samples combined across $m = 10$ multiply imputed datasets. Included predictors had a statistically significant permutation test z-score for at least one of the multiply imputed training datasets. Additional predictors that did not achieve statistical significance via the permutation test, but were used in the `RFRE.PO` algorithm in some form, are not shown. For analysis, we created three history variables, $H_f(t), H_p^{(2)}(t)$ and $H_p^{(3)}(t)$. $H_f(t)$ was an indicator if the subject had experienced a severe exacerbation on study time prior to time t. $H_p^{(2)}(t)$ was assumed to take a categorical form based on whether the participant had an exacerbation in the past zero to 31 days ($H_p^{(2)}(t) = 4$), 31-92 days ($H_p^{(2)}(t) = 3$), 93-182 days ($H_p^{(2)}(t) = 2$), 183 days to 365 days ($H_p^{(2)}(t) = 1$), or finally, if that time was more than 365 days or never ($H_p^{(2)}(t) = 0$). $H_p^{(3)}(t)$ is the estimated 30-day event rate using all follow-up prior to time t.

Covariate	p
Categorical $H_{p,i}^{(2)}(t) = 0$ (versus Categorical $H_{p,i}^{(2)}(t) = 1, 2, 3, 4$)	0.06
Categorical $H_{p,i}^{(2)}(t) = 1$ (versus Categorical $H_{p,i}^{(2)}(t) = 0, 2, 3, 4$)	<**0.04**
Categorical $H_{p,i}^{(2)}(t) = 4$ (versus Categorical $H_{p,i}^{(2)}(t) = 0, 1, 2, 3$)	<**0.01**
Categorical $H_{p,i}^{(2)}(t)$	<**0.01**
$H_{p,i}^{(3)}(t)$	<**0.01**
$H_{f,i}(t)$	<**0.01**
Respiratory Hospitalization in the Year before Study Start	**0.01**
Corticosteroid Use in the Year before Study Start	<**0.01**
Gender	0.99
Smoking at Baseline	0.98
General Health Score at Time t	<**0.01**
Standardized General Health Z-Score at Time t	<**0.01**
Pain Score at Time t	**0.03**
Standardized Pain Score at Time t	**0.05**
SGRQ Symptom Score at Time t	<**0.01**
Seen a Mental Health Professional for Anxiety/Depression at Time t	1.00
Anti-Coagulant Use at Time t	1.00
Inhaled Corticosteroids, Long-Acting Muscarinic, and Long-Acting Bronchodilator Agent at Time t	0.99
Long-Acting Muscarinic Use at Time t	1.00
Treatment Group	**0.02**
Leukocytosis at Time t	**0.02**
Oxygen Use at Time t	1.00
Bronchiectasis Diagnosis at Time t	<**0.01**
Phlegm at Time t	1.00

FEV_1 is forced expiratory volume in one second and FVC is forced vital capacity in liters, clinical measures of lung function.

SGRQ is the Saint George's Respiratory Questionnaire, a patient reported outcome measuring quality of life.

Table S8: Comparison of the C-statistics and mean-squared errors (MSE) of models fit to the validation set of the Azithromycin in the Prevention of COPD cohort. The Wald forward selection model and the model corresponding to clinical input from Xia et al. (2020) are fit using the modified XMT approach. The forward selection model includes predictors through automated forward selection requiring $p < 0.05$ for entry into the model. Missing values in the forward selection model and `RFRE.PO` for time since most recent exacerbation are multiply imputed using the coding scheme described in Section 8. C-statistic and mean-squared error estimates and standard errors for all methods and models are averaged across results from multiply imputed data with Rubin's rule applied to bootstrapped ($b = 100$) standard errors.

Model Name	C-Statistic		MSE	
	Mean	SD	Mean	SD
`RFRE.PO`	0.613	0.009	0.242	0.003
m-XMT (Clinical Input)	0.565	0.009	0.252	0.002
m-XMT (Forward Selection)	0.570	0.009	0.264	0.004

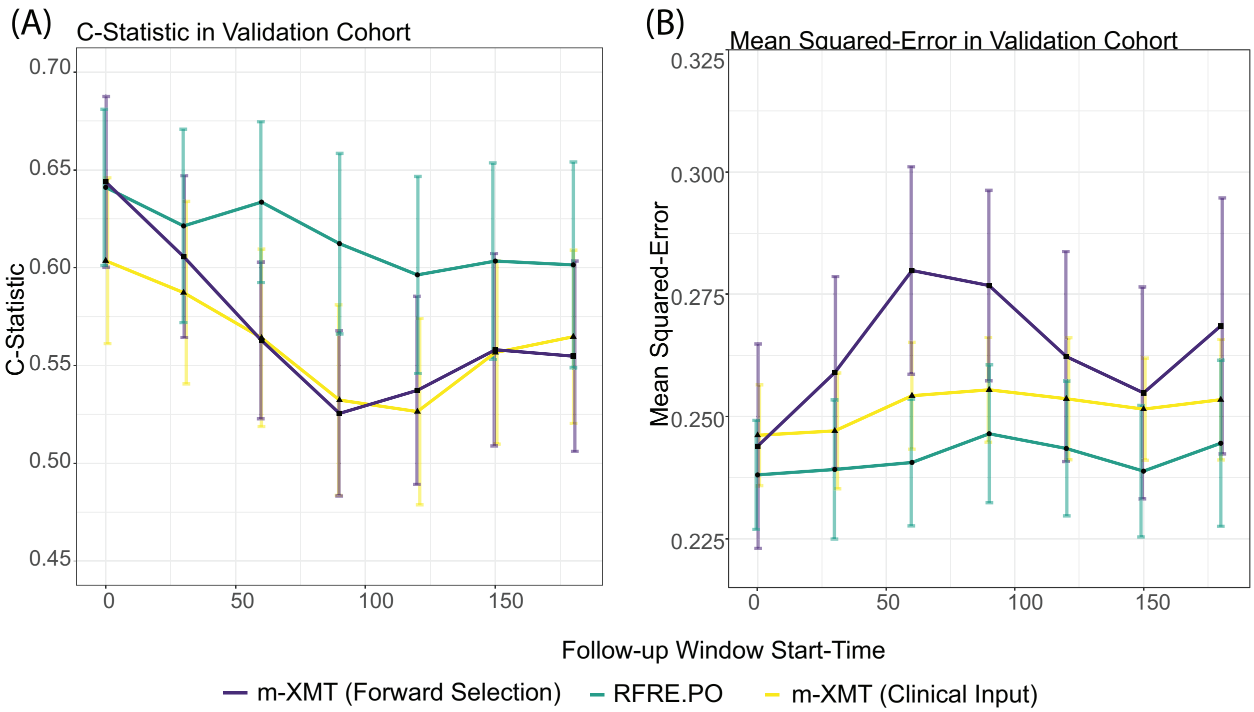

Figure S3: A panel graphic displaying time-dependent C-statistics (Panel A) and MSEs (Panel B) in the validation dataset of the Azithromycin for the Prevention of COPD study, calculated at each follow-up window $t \in \{0, 30, 60, 90, 120, 150, 180\}$ for $\tau = 180$; 95% confidence intervals also displayed. Compared to the other algorithms, `RFRE.PO` has fairly stable and attractive performance metrics across the various follow-up windows, likely because interactions with the various predictors over time are naturally incorporated if useful for prediction.

References

Xia, M., Murray, S., and Tayob, N. (2020). Regression analysis of recurrent-event-free time from multiple follow-up windows. *Statistics in Medicine*, 39:1–15.