Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Global and Episode-Specific Prediction of Recurrent Events Using Longitudinal Health Informatics Data

Yifei Sun, Sy Han Chiou & Chiung-Yu Huang

To cite this article: Yifei Sun, Sy Han Chiou & Chiung-Yu Huang (2025) Global and Episode-Specific Prediction of Recurrent Events Using Longitudinal Health Informatics Data, Journal of the American Statistical Association, 120:552, 2015-2027, DOI: 10.1080/01621459.2025.2497569

To link to this article: https://doi.org/10.1080/01621459.2025.2497569

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 120, NO. 552, 2015–2027: Applications and Case Studies
https://doi.org/10.1080/01621459.2025.2497569

Global and Episode-Specific Prediction of Recurrent Events Using Longitudinal Health Informatics Data

Yifei Sun[a], Sy Han Chiou[b], and Chiung-Yu Huang[c]

[a]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY; [b]Department of Statistics and Data Science, Southern Methodist University, Dallas, TX; [c]Department of Epidemiology & Biostatistics, University of California at San Francisco, San Francisco, CA

ABSTRACT

Accurate prediction of recurrent clinical events is crucial for effective management of chronic conditions such as cancer and cardiovascular disease. In recent years, longitudinal health informatics databases, which routinely collect data on repeated clinical events, have been increasingly used to construct risk prediction models. We introduce a novel nonparametric framework to predict recurrent events on a gap time scale using survival tree ensembles. Our framework incorporates two predictive modeling strategies: episode-specific model and global model. These models avoid strong assumptions on how future event risk depends on previous event history and other predictors, making them a promising alternative to Cox-type models. Additional complexities in tree-based prediction for recurrent events include induced informative censoring of gap times and inter-event correlations. We develop algorithms to address these issues through the use of inverse probability of censoring weighting and modified resampling procedures. Applied to SEER-Medicare data to predict repeated hospitalizations for breast cancer patients, our models showed superior performance. In particular, borrowing information across events via global models substantially improved prediction accuracy for later hospitalizations. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received January 2024
Accepted April 2025

KEYWORDS
Dynamic risk prediction; Ensemble; Gap time; Medical records; Random forest

1. Introduction

Recurrent events, which are common in chronic diseases, are characterized by multiple occurrences over time. Accurate prediction of recurrent events enables early identification of high-risk patients for intensive monitoring. This shift from reactive to preventive and personalized healthcare has the potential to improve patients' quality of life and survival. Baseline predictors, such as demographics and disease severity, are commonly used to predict single events, but they may have limited predictive power for subsequent events. To improve the accuracy of recurrent event prediction, it is important to use up-to-date predictor information, such as historical data on events and other time-varying factors. For example, patients with a history of major depressive disorder are more likely to experience a subsequent episode of depression than those without prior occurrences, even if they have similar baseline characteristics (Nuggerud-Galeas et al. 2020). As another example, patients with a history of breast cancer (BC) are at increased risk of developing a second BC compared to the general population (Pedersen et al. 2022).

The rapid growth of longitudinal health informatics data, including electronic health records and medical claims, has opened up new opportunities for developing risk prediction tools (Beam and Kohane 2018). These data repositories routinely document the timing of recurring clinical events and relevant patient health information, allowing researchers to construct prediction models for recurrent events. Among these valuable data sources is the SEER-Medicare linked database (Warren et al. 2002; Enewold et al. 2020), which combines cancer registry data with Medicare claims. Our goal is to leverage the rich information in SEER-Medicare to develop prediction models for recurrent hospitalizations in BC patients. Recurrent hospitalizations, often triggered by complications arising from post-treatment side effects, coexisting conditions, or overall health decline, pose a significant challenge in BC management. Addressing this concern necessitates the development of dynamic predictive models that accurately identify patients at high risk of recurrent health problems.

Event risk prediction usually relies on modeling the intensity function of the event process, which characterizes the instantaneous risk of an event occurring given the history of events and covariates. Cox-type regression models are a popular class of intensity models that assume multiplicative covariate effects. These models are typically used with one of two time scales: total elapsed time or gap time (time from the preceding event). Additionally, they can be adapted to incorporate stratification, allowing for different baseline intensity functions and regression coefficients for different event episodes. For example, Andersen and Gill (1982) specified a proportional intensity model with common covariate effects across all episodes and used the elapsed time scale. Prentice, Williams, and Peterson (1981) considered stratified proportional intensity models that allow the baseline intensity function and covariate effects to depend on the number of preceding failures, making them stratum-

CONTACT Yifei Sun ys3072@cumc.columbia.edu Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

specific. Their models can accommodate both the total elapsed time and the gap time. Chang and Hsiung (1994) considered both time scales and modeled covariate effects shared across episodes. Finally, Chang and Wang (1999) developed a model based on the gap time scale where the baseline intensity function is stratified by the number of preceding events, and allowed the covariate effects to be global or episode-specific.

Existing methods for recurrent events typically use Cox-type models to facilitate statistical inference. However, the assumption of multiplicative covariate effects can be restrictive, and correctly specifying the complex relationships between future events and accumulating historical information is challenging. In cases of model misspecification, the accuracy of predicted risk may be compromised. Nonparametric methods, such as decision tree learning, emerge as promising alternatives. For a single event, tree-based methods have demonstrated superior performance when the true model is complicated and deviates from Cox-type models. Building upon the concept of classification and regression trees (Breiman et al. 1984), survival trees use baseline predictors as input and estimate survival probability conditioned on these predictors (Gordon and Olshen 1985; Segal 1988; Davis and Anderson 1989; LeBlanc and Crowley 1992, 1993; Molinaro, Dudoit, and Van der Laan 2004; Steingrimsson et al. 2016). As a single tree is not stable, ensemble methods have been introduced (Hothorn et al. 2004, 2006; Ishwaran et al. 2008; Zhu and Kosorok 2012; Steingrimsson, Diao, and Strawderman 2019). Tree-based methods to predict recurrent events remain relatively underexplored. Sparapani et al. (2020) proposed Bayesian additive regression trees for nonparametric modeling of recurrent events, employing total elapsed time as the time scale without stratification.

We introduce survival tree ensembles for dynamically predicting recurrent events on the gap time scale. We propose two types of model: episode-specific prediction, where a separate model is built for each event, and global prediction, where one model is built for multiple events. Episode-specific models capture the unique characteristics of each event, providing full stratification and detailed insights into the pattern for each event. Global models offer a data-driven approach to determine the level of stratification. The proposed methods enjoy several advantages over existing semiparametric approaches. First, they can accommodate an increasing number of predictors for later events, thereby allowing for more effective use of accumulating information over time. Second, the global models leverage information across multiple episodes, which can improve prediction accuracy while circumventing the need to select stratification levels. Lastly, similar to other tree-based methods, the proposed methods require minimal assumptions; they do not rely on proportional intensities or hazards, nor do they require the specification of complex dependency structures among recurrent events. The transition from a single event to the recurrent event setting brings several challenges, including induced dependent censoring on the gap time scale and correlation among event episodes. We address these by applying the inverse probability of censoring weighting (IPCW) in tree-building and ensemble procedures, and by employing modified resampling techniques, such as within-cluster resampling (Hoffman, Sen, and Weinberg 2001; Follmann, Proschan, and Leifer 2003), in our global models.

2. Data Structure and Model Setup

2.1. Episode-Specific Models versus Global Models

We focus on predicting the risk of q recurrent events, where q is a pre-specified positive integer. For $j = 1, \ldots, q$, denote by T^j the time at which the jth recurrent event occurs and assume T^j is continuous. The outcome of interest is the gap time between consecutive events, that is, $G^{j+1} = T^{j+1} - T^j$ for $j = 0, 1, \ldots, q-1$, with $T^0 = 0$ representing time zero. Note that an individual may experience fewer than q events even if they are not lost to follow-up. If an individual has only q' events, where $q' < q$, we set $G^{q'+1} = \cdots = G^q = \infty$ and $T^{q'+1} = \cdots = T^q = \infty$. Denote by $\boldsymbol{W}$ a p-dimensional vector of baseline predictors. In the SEER-Medicare data, hospitalizations are recurrent events and the diagnosis of BC defines time zero. The baseline predictors include various demographic and clinical variables collected at the time of diagnosis. Moreover, information on previous hospitalizations can be used to predict future hospitalizations.

In this article, we consider two types of prediction models:

(I) Episode-specific prediction: This approach builds a separate model for each event. To predict the $(j+1)$th event $(1 \le j \le q-1)$, we reset the clock to zero at T^j. Predictors for G^{j+1}, denoted by $\boldsymbol{X}^j$, can include $\boldsymbol{W}$ and features derived from event history up to time T^j. For example, one may define $\boldsymbol{X}^j = (\boldsymbol{W}^\top, G^1, \ldots, G^j)^\top$ to incorporate all available historical information. For $t \ge 0$, the target survival function for the $(j+1)$th event is defined as

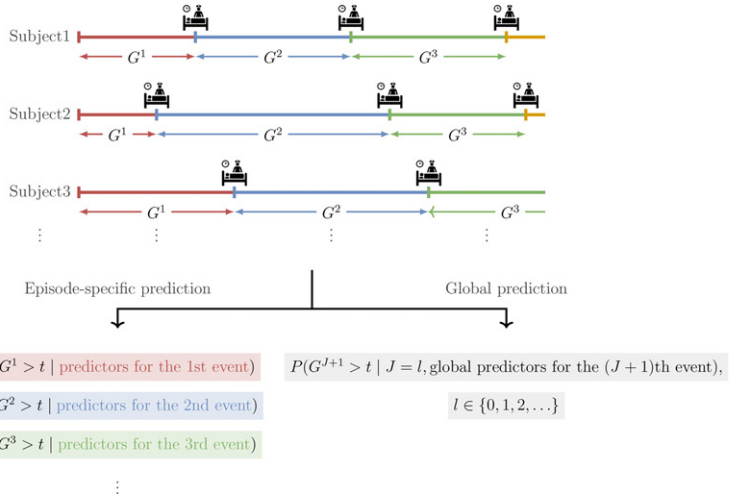
$$S^{j+1}(t \mid \boldsymbol{x}^j) \stackrel{\text{def}}{=} P(G^{j+1} > t \mid \boldsymbol{X}^j = \boldsymbol{x}^j).$$

(II) Global prediction: The global model uses a single overarching model to predict multiple events. Let J be a random variable indicating the number of events that have occurred thus far. Its support, $\mathcal{J} \subseteq \{0, 1, 2, \ldots, q-1\}$, includes more than one value. We define J to be independent of other variables $\boldsymbol{W}, G^1, \ldots, G^q$. The target survival function is defined as

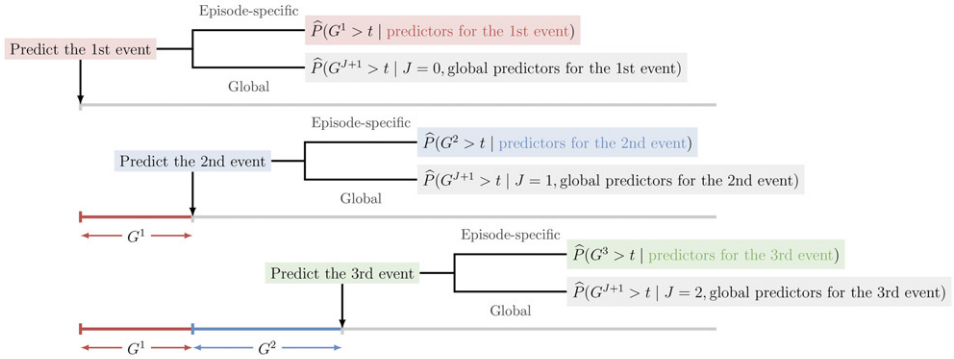
$$S(t \mid \boldsymbol{z}^l) \stackrel{\text{def}}{=} P(G^{J+1} > t \mid \boldsymbol{Z}^J = \boldsymbol{z}^l), \quad l \in \mathcal{J},$$

where G^{J+1} denotes the next gap time and takes the value of G^{j+1} when $J = j$; $\boldsymbol{Z}^J$ denotes predictors observed up to the time of the Jth event, with its dimension remaining constant regardless of the value of J. To accommodate episode-specific distributions, $\boldsymbol{Z}^J$ includes J as one of its elements. Further details on the specification of $\boldsymbol{Z}^J$ are provided below.

Figure 1 illustrates the episode-specific and global models and their application in predicting future events. Upon the occurrence of a preceding event, both models can be used to predict the next event. The difference between episode-specific and global models parallels the contrast between stratified regression models and an overall model that includes stratification factors as covariates. In conventional regression analysis, associations between covariates and outcomes can be evaluated within each stratum (e.g., gender), resulting in stratum-specific models (e.g., separate models for males and females). Likewise, an episode-specific model predicts outcomes separately for each episode,



(a) Patient records with recurrent hospitalizations and two types of modeling strategies



(b) Risk prediction for a new subject ($\widehat{P}$ denotes the estimated probability)

Figure 1. Data structure and dynamic prediction of recurrent events.

allowing the influence of predictors to vary across episodes. Here, the episode number acts like the stratification variable. In contrast, an overall model regresses the outcome on both the covariates and the stratification variable (e.g., gender), possibly including interaction terms to capture variations in covariate effects across strata. In parallel, a global model treats the episode number J as an additional predictor. The nonparametric target survival function accommodates potential interactions between episode number and other predictors, capturing overarching trends while allowing for variation in predictor effects across episodes.

We consider two approaches for setting up global models: (a) a simple, *partially global model* that borrows information across a subset of events, and (b) an *enhanced global model* that borrows information across all events.

The partially global model uses the most recent events as predictors and predicts the $(m+1)$th and subsequent events. The first m events can be predicted separately using episode-specific models. In practice, m can be treated as a tuning parameter and selected using cross-validation. Section 6 provides a detailed explanation of how this was implemented in the data analysis. The upper panel of Figure 2 illustrates how multiple episodes contribute to this model for $m = 0, 1, \ldots, q-2$. For a given m, let $\mathcal{J} = \{m, m+1, \ldots, q-1\}$ be the support of J. For $k = 0, 1, \ldots, m$, we define G^{J-k+1} as the $(k-1)$th gap time when counting backward from J, and G^{J-k+1} takes the value of G^{j-k+1} when $J = j$. Similarly, T^J is assigned the value of T^j when $J = j$. The global model predicts G^{J+1} using predictors $\boldsymbol{Z}^J = (\boldsymbol{W}^\top, T^J, H(J,m)^\top, J)^\top$, where $H(J,m) = (G^J, G^{J-1}, \ldots, G^{J-m+1})^\top$ denotes the history of the m preceding gap times. For $m \le l \le q-1$, $\{(G^{J+1}, \boldsymbol{W}^\top, T^J, H(J,m)^\top)^\top \mid J = l\}$ is distributed identically to $(G^{l+1}, \boldsymbol{W}^\top, T^l, H(l,m)^\top)^\top$, where $H(l,m) = (G^l, G^{l-1}, \ldots, G^{l-m+1})^\top$. We write $\boldsymbol{z}^l = (\boldsymbol{w}^\top, t^l, g^l, \ldots, g^{l-m+1}, l)^\top$, and the global target function is expressed as

$$\begin{aligned} S(t \mid \boldsymbol{z}^l) &\stackrel{\text{def}}{=} P(G^{J+1} > t \mid \boldsymbol{Z}^J = \boldsymbol{z}^l) \\ &= P(G^{l+1} > t \mid \boldsymbol{W} = \boldsymbol{w}, T^l = t^l, \\ &\quad G^l = g^l, \ldots, G^{l-m+1} = g^{l-m+1}), \quad m \le l \le q-1. \end{aligned} \tag{1}$$

The expression integrates the distributions of the $(m+1)$th and subsequent events into one function $S(t \mid \boldsymbol{z}^l)$. To estimate $S(t \mid \boldsymbol{z}^l)$, we can use multiple episodes. As shown in Figure 2, a larger m allows the inclusion of more historical information as predictors, albeit at the cost of fewer episodes for model training.

The partially global model ignores early events as predictors and does not exploit all available observations. To overcome this limitation, we propose an *enhanced global model* that extends the support of J to $\mathcal{J} = \{0, 1, \ldots, q-1\}$. An illustration of

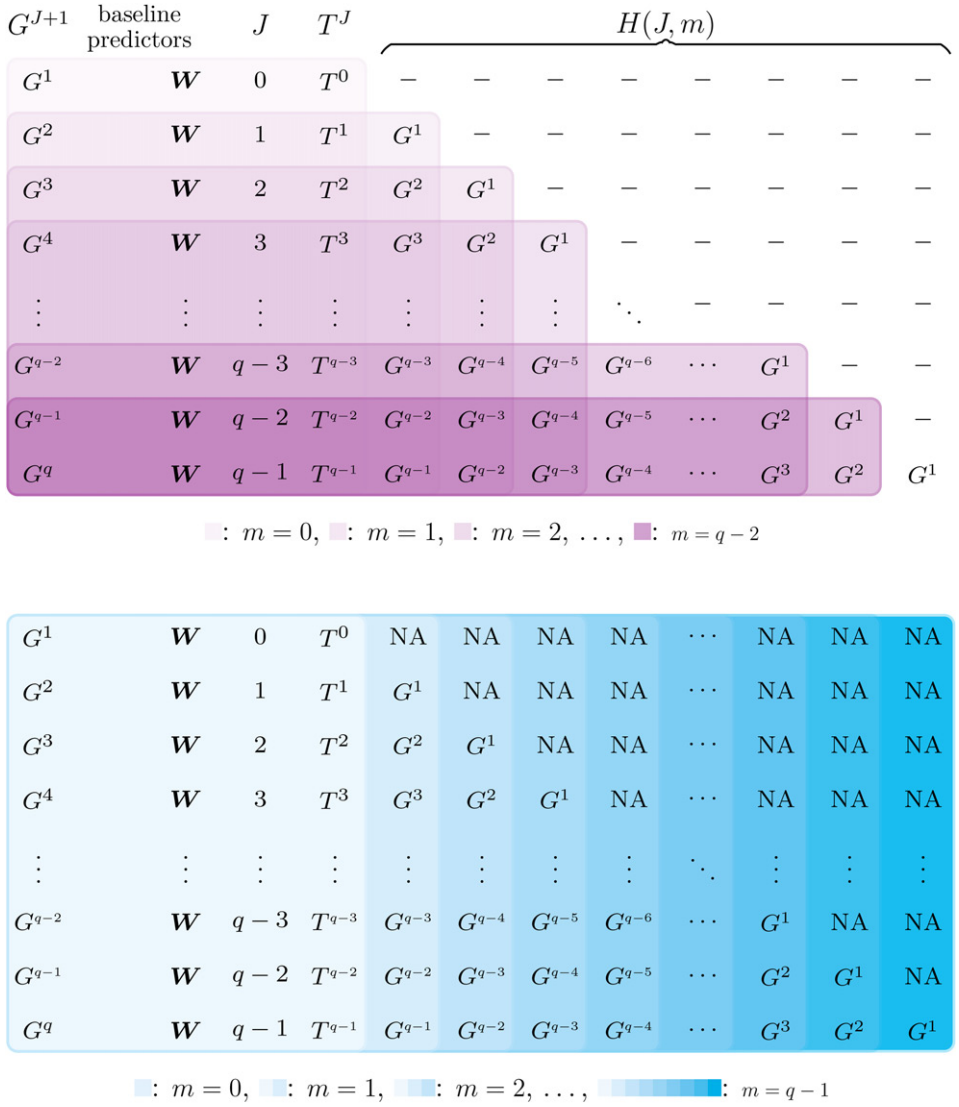


Figure 2. Illustration of outcome G^{J+1} and predictors $\boldsymbol{Z}^J = (\boldsymbol{W}^\top, T^J, H(J,m)^\top, J)^\top$ in the augmented training data for global models. Top panel: The layers indicate observations used in the partially global model when m previous gap times are used as predictors. Increasing m expands the predictor dimension horizontally while reducing the number of available observations vertically. Bottom panel: The layers indicate the observations used in the enhanced global models when up to m previous gap times are used as predictors.

how repeated observations contribute to the enhanced global model is given in the lower panel of Figure 2. We generalize the definition of $H(l,m)$ to accommodate the case where $l < m$. As with the previous definition, $H(l,m)$ is an m-dimensional vector with its kth element being G^{l-k+1}. For $l-k+1 \le 0$, G^{l-k+1} has not been defined, and thus we set $G^{l-k+1} \equiv$ NA. Here, "NA" signifies "not available" and represents a non-numeric value. Note that NA serves as a placeholder rather than missing data, because the $(l-k+1)$th gap time does not exist when $l-k+1 \le 0$. Therefore, when $l < m$, $H(l,m)$ is represented as $H(l,m) = (G^l, \ldots, G^1, \underbrace{\text{NA}, \ldots, \text{NA}}_{m-l})^\top$, where the last $m-l$ elements are NA. Moreover, we define $H(J,m)$, T^J, G^{J+1} to take the values of $H(l,m)$, T^l, G^{l+1}, respectively, when $J = l$. Then $\{(G^{J+1}, \boldsymbol{W}^\top, T^J, H(J,m)^\top)^\top \mid J = l\}$ is distributed identically to $(G^{l+1}, \boldsymbol{W}^\top, T^l, H(l,m)^\top)^\top$ for $0 \le l \le q-1$. For $0 \le l \le m-1$, we write $\boldsymbol{z}^l = (\boldsymbol{w}^\top, t^l, g^l, \ldots, g^1, \underbrace{\text{NA}, \ldots, \text{NA}}_{m-l}, l)^\top$. The target function is expressed as

$$
\begin{aligned}
S(t \mid \boldsymbol{z}^l) &\stackrel{\text{def}}{=} P(G^{J+1} > t \mid \boldsymbol{Z}^J = \boldsymbol{z}^l) \\
&= \begin{cases} P(G^{l+1} > t \mid \boldsymbol{W} = \boldsymbol{w}, T^l = t^l, G^l = g^l, \ldots, G^1 = g^1), \\ \quad \text{if } 0 \le l \le m-1, \\ P(G^{l+1} > t \mid \boldsymbol{W} = \boldsymbol{w}, T^l = t^l, G^l = g^l, \ldots, G^{l-m+1} \\ \quad = g^{l-m+1}), \quad \text{if } m \le l \le q-1. \end{cases}
\end{aligned} \tag{2}
$$

In this model, m determines the amount of historical information as predictors. To fully use historical information, we can set $m = q-1$.

When j is large and the follow-up period is short, the episode-specific model for the $(j+1)$th event can be unstable due to the limited number of subjects experiencing more than j events. Global models, on the other hand, can leverage information from multiple episodes, making them an attractive option when there is a high degree of similarity in event-predictor relationships across episodes. Intuitively, when events are highly correlated with the same set of predictors, using multiple episodes

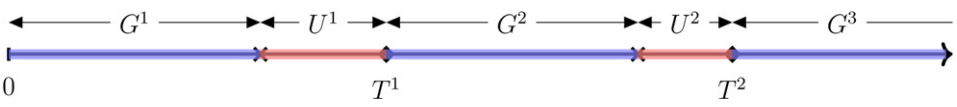


Figure 3. Illustration of data structure accounting for length of stay.

may increase the chance of identifying important predictors. When event-predictor relationships vary substantially across events, global models may offer less notable improvements.

2.2. Additional Considerations and Observed Data

While the preceding discussion assumes a negligible length of stay (LoS) in the hospital, it is important to recognize that patients are not at risk of rehospitalization when hospitalized. In scenarios where the LoS is notable, it is more practical to predict the next hospital admission at discharge rather than admission and include LoS as a predictor. To incorporate LoS, the notation needs to be redefined. Let G^1 represent the time of the first hospital admission. For $j \geq 2$, we redefine G^j as the time between the discharge of the $(j-1)$th hospitalization and the subsequent hospital admission. Each gap time is followed by a LoS, denoted as U^j for the jth hospitalization, $j = 1, \ldots, q$. Figure 3 illustrates the data structure that accounts for LoS. For $j \geq 1$, we define $T^j = \sum_{k=1}^{j}(G^k + U^k)$, which represents the time of discharge from the jth hospitalization, and set $T^0 = 0$. Prediction of G^{j+1} is made at T^j. The notation introduced in Section 2.1 represents a special case where $U^j = 0$. Variables in the global model, such as G^{J+1} and T^J, retain their original definitions, but are now based on the updated $\{G^j, T^j, j = 1, \ldots, q\}$ to account for LoS.

Our discussion has primarily focused on using the history of gap times as predictors. Other historical information, such as the LoS, can also be used as predictors. In episode-specific precision, we can simply expand the predictor vector as $\boldsymbol{X}^j = (\boldsymbol{W}^\top, G^1, U^1, G^2, U^2, \ldots, G^j, U^j)^\top$. In global prediction, we can define $H(l, m)$ as $(G^l, U^l, \ldots, G^{l-m+1}, U^{l-m+1})^\top$ when $l \geq m$, and as $(G^l, U^l, \ldots, G^1, U^1, \underbrace{\text{NA}, \ldots, \text{NA}}_{2(m-l)})^\top$ when $l < m$. Then $S(t \mid \boldsymbol{z}^l)$ in (1) and (2) can be easily extended to condition on both the previous gap times and the corresponding LoS. Furthermore, summary measures derived from event history, such as the percent change in the length of gap times $(G^j - G^{j-1})/G^{j-1} \times 100\%$, can also serve as predictors. Intuitively, a larger percent increase in gap times might suggest that a patient's health is improving and indicate a lower risk of future events.

In practice, event times are usually subject to censoring. Let C denote the time to the end of follow-up. We assume that C is continuous and independent of $(G^1, U^1, \ldots, G^q, U^q)^\top$ given $\boldsymbol{W}$. For $j \geq 1$, the observed total time to the jth hospital discharge is defined as $Y^j = \min(T^j, C)$, with the event indicator $\delta^j = I(T^j \leq C)$. For individuals who remain under observation at T^j (i.e., $\delta^j = 1$), predictors for G^{j+1} are completely observed; moreover, we observe a possibly censored gap time, $\widetilde{G}^{j+1} = \min(G^{j+1}, C - T^j)$, and the corresponding event indicator is defined as $\Delta^{j+1} = I(T^j + G^{j+1} \leq C)$. For individuals no longer under observation at T^j, we set $\Delta^{j+1} = 0$ and $\widetilde{G}^{j+1} = 0$.

3. Episode-Specific Prediction

3.1. Survival Tree Ensembles

Episode-specific models predict G^{j+1} using $\boldsymbol{X}^j$ for a given j $(1 \leq j \leq q-1)$, where $\boldsymbol{X}^j$ is observed when $\delta^j = 1$. The training data are $\{\widetilde{G}_i^{j+1}, \Delta_i^{j+1}, \boldsymbol{X}_i^j \delta_i^j, \delta_i^j, C_i, \boldsymbol{W}_i; i = 1, \ldots, n\}$. When incorporating LoS, define the predictor space $\mathcal{X}^j = \{\boldsymbol{x}^j = (\boldsymbol{w}^\top, g^1, u^1, \ldots, g^j, u^j)^\top : \boldsymbol{w} \in \mathcal{W}, \sum_{k=1}^{j}(g^k + u^k) \leq t_{\max}, g^k \geq 0, u^k \geq 0, k = 1, \ldots, j\}$, where $\mathcal{W}$ is the sample space of $\boldsymbol{W}$, and $t_{\max}$ is a pre-specified constant. The constraint $T^j = \sum_{k=1}^{j}(G^k + U^k) \leq t_{\max}$ is imposed to ensure stable estimation of $S^{j+1}(t \mid \boldsymbol{x}^j)$ for $\boldsymbol{x}^j \in \mathcal{X}^j$. We denote the conditional hazard of G^{j+1} given $\boldsymbol{X}^j = \boldsymbol{x}^j$ by $\lambda^{j+1}(t \mid \boldsymbol{x}^j)$, and we have $S^{j+1}(t \mid \boldsymbol{x}^j) = \exp\{-\int_0^t \lambda^{j+1}(u \mid \boldsymbol{x}^j)du\}$. Under the independent censoring assumption, hazard is expressed as a ratio of conditional expectations involving only observed variables:

$$\lambda^{j+1}(t \mid \boldsymbol{x}^j)dt = \frac{E\{dN^{j+1}(t) \mid \boldsymbol{X}^j = \boldsymbol{x}^j, \delta^j = 1\}}{E\{I(\widetilde{G}^{j+1} \geq t) \mid \boldsymbol{X}^j = \boldsymbol{x}^j, \delta^j = 1\}}, \quad (3)$$

where $N^{j+1}(t) = \Delta^{j+1} I(\widetilde{G}^{j+1} \leq t)$. Based on (3), these conditional expectations and thus $S^{j+1}(t \mid \boldsymbol{x}^j)$ can be estimated using the subsample with observed predictors.

Decision trees segment the predictor space using hierarchical rules. They are typically constructed using a recursive partitioning algorithm, where we start with the root node representing $\mathcal{X}^j$, split each node into two child nodes by comparing a predictor with a threshold (the split is optimized based on some criterion) and repeat (Breiman et al. 1984). This process divides $\mathcal{X}^j$ into multiple subsets called terminal nodes, and the subsets formed during the intermediate stages are called internal nodes. Let $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_M\}$ be the set of terminal nodes such that $\tau_1 \cup \tau_2 \cup \cdots \cup \tau_M = \mathcal{X}^j$ and $\tau_k \cap \tau_{k'} = \emptyset$ for $k \neq k'$. Within each terminal node, observations are assigned the same target value, estimated based on the training data within the node. We define the partition function $\ell_{\mathcal{T}} : \mathcal{X}^j \mapsto \mathcal{T}$ that maps $\boldsymbol{x}^j$ to the terminal node that contains it.

Heuristically, if the terminal node $l_{\mathcal{T}}(\boldsymbol{x}^j)$ defines a small neighborhood of $\boldsymbol{x}^j$, one may approximate the terms in (3) with their within-node counterparts, $E\{dN^{j+1}(t) \mid \boldsymbol{X}^j \in \ell_{\mathcal{T}}(\boldsymbol{x}^j), \delta^j = 1\}$ and $E\{I(\widetilde{G}^{j+1} \geq t) \mid \boldsymbol{X}^j \in \ell_{\mathcal{T}}(\boldsymbol{x}^j), \delta^j = 1\}$. This leads to a simple partition-based estimator for $\lambda^{j+1}(t \mid \boldsymbol{x}^j)dt$: $\frac{\widehat{E}\{dN^{j+1}(t) | \boldsymbol{X}^j \in \ell_{\mathcal{T}}(\boldsymbol{x}^j), \delta^j = 1\}}{\widehat{E}\{I(\widetilde{G}^{j+1} \geq t) | \boldsymbol{X}^j \in \ell_{\mathcal{T}}(\boldsymbol{x}^j), \delta^j = 1\}}$, where $\widehat{E}$ denotes the empirical average evaluated using the training data. However, the estimator may not provide an accurate estimate of the within-node hazard when the node $l_{\mathcal{T}}(\boldsymbol{x}^j)$ is "large" and contains a wide range of predictor values. We define the within-node hazard for a given node τ (which can either be terminal or internal) as $\lambda^{j+1}(t \mid \tau)dt \stackrel{\text{def}}{=} P(G^{j+1} \in [t, t+dt) \mid G^{j+1} \geq t, \boldsymbol{X}^j \in \tau)$. As shown in the supplementary materials, the ratio $\frac{\widehat{E}\{dN^{j+1}(t) | \boldsymbol{X}^j \in \tau, \delta^j = 1\}}{\widehat{E}\{I(\widetilde{G}^{j+1} \geq t) | \boldsymbol{X}^j \in \tau, \delta^j = 1\}}$ does not estimate $\lambda^{j+1}(t \mid \tau)dt$. This is due to induced dependent censoring, that is, correlation between censoring time $C - T^j$ and the

gap time G^{j+1} within τ, even when censoring is independent on the total elapsed time scale. Inaccurate estimates within nodes, especially during the early stages of tree construction when nodes are large, may lead to suboptimal splits and decreased prediction accuracy. To address induced dependent censoring, we apply the IPCW. This is motivated by the following equation:

$$\lambda^{j+1}(t \mid \tau)dt = \frac{E\{I(\boldsymbol{X}^j \in \tau)dN^{j+1}(t)/S_C(Y^j + t \mid \boldsymbol{W})\}}{E\{I(\boldsymbol{X}^j \in \tau, \widetilde{G}^{j+1} \geq t)/S_C(Y^j + t \mid \boldsymbol{W})\}}, \quad (4)$$

where $S_C(u \mid \boldsymbol{W}) = P(C > u \mid \boldsymbol{W})$. Compared to (3), $\delta^j = 1$ is not present here because both $dN^{j+1}(t) = 1$ and $I(\widetilde{G}^{j+1} \geq t) = 1$ imply $\delta^j = 1$. In practice, nonparametric methods, such as random forests, can be used to estimate $S_C(\cdot \mid \boldsymbol{W})$, and we denote the estimator by $\widehat{S}_C(\cdot \mid \boldsymbol{W})$. Unlike conventional survival analysis, where the censoring time is not observed when the event occurs, the follow-up time C is always observed in our setting. Therefore, estimation of $S_C(\cdot \mid \boldsymbol{W})$ can be performed using methods for complete data. When censoring is completely random, the empirical marginal survival probability of C can be used.

Due to the instability of a single tree, we focus on ensemble models in line with random forests (Breiman 2001). We generate B datasets (e.g., $B = 500$) using bootstrap or subsampling among all subjects. For each resampled dataset, a tree is constructed using recursive partitioning guided by a concordance-based criterion described in Section 3.2. At each split, a random subset of predictors is selected as candidate splitting variables. For the bth tree ($b = 1, 2, \ldots, B$), we denote the terminal node containing $\boldsymbol{x}^j$ by $\ell_b(\boldsymbol{x}^j)$. According to (4), the bth tree provides the following estimating equation for $\lambda^{j+1}(t \mid \boldsymbol{x}^j)dt$: $\sum_{i=1}^n w_{bi} I(\boldsymbol{X}_i^j \in \ell_b(\boldsymbol{x}^j))\widehat{S}_C(Y_i^j + t \mid \boldsymbol{W}_i)^{-1}\{dN_i^{j+1}(t) - I(\widetilde{G}_i^{j+1} \geq t)\lambda^{j+1}(t \mid \boldsymbol{x}^j)dt\} = 0$, where w_{bi} denotes the frequency of the ith subject in the bth resampled dataset. Averaging these estimating equations across all trees, we obtain the ensemble estimating equation: $\sum_{i=1}^n w_{ij}(\boldsymbol{x}^j, t)\{dN_i^{j+1}(t) - I(\widetilde{G}_i^{j+1} \geq t)\lambda^{j+1}(t \mid \boldsymbol{x}^j)dt\} = 0$, where $w_{ij}(\boldsymbol{x}^j, t) = B^{-1}\sum_{b=1}^B w_{bi}\delta_i^j I(\boldsymbol{X}_i^j \in \ell_b(\boldsymbol{x}^j))/\widehat{S}_C(Y_i^j + t \mid \boldsymbol{W}_i)$. Given $S^{j+1}(t \mid \boldsymbol{x}^j) = \exp\{-\int_0^t \lambda^{j+1}(u \mid \boldsymbol{x}^j)du\}$, we propose to estimate the target survival function by:

$$\widehat{S}^{j+1}(t \mid \boldsymbol{x}^j) = \exp\left\{-\int_0^t \frac{\sum_{i=1}^n w_{ij}(\boldsymbol{x}^j, u)dN_i^{j+1}(u)}{\sum_{i=1}^n w_{ij}(\boldsymbol{x}^j, u)I(\widetilde{G}_i^{j+1} \geq u)}\right\}. \quad (5)$$

The weight function, $w_{ij}(\boldsymbol{x}^j, u)$, assigns higher weights to observations frequently classified within terminal nodes containing $\boldsymbol{x}^j$ and to those with a later occurrence of the jth event. This weighting scheme not only defines a forest-based neighborhood around $\boldsymbol{x}^j$, but also takes into account the impact of incomplete follow-up. Equation (5) introduces a novel ensemble estimator that accounts for the induced dependent censoring specific to the gap time framework, a complexity not addressed by existing machine learning methods.

3.2. A Concordance Metric for Model Accuracy and Splitting

In this section, we introduce a time-dependent concordance metric as both a measure of prediction performance and a criterion for splitting. We denote $g(t, \boldsymbol{X}^j)$ as a score that quantifies the risk of $G^{j+1} \leq t$. For example, $g(t, \boldsymbol{X}^j)$ can be the estimated probability of $G^{j+1} \leq t$ given $\boldsymbol{X}^j$. The concordance between $g(t, \boldsymbol{X}^j)$ and $G^{j+1} \leq t$ is defined as

$$\begin{aligned}\mathrm{CON}_t^{j+1}(g) &= P\{g(t, \boldsymbol{X}_1^j) < g(t, \boldsymbol{X}_2^j) \mid G_1^{j+1} > t \\ &\quad \geq G_2^{j+1}, \boldsymbol{X}_1^j \in \mathcal{X}^j, \boldsymbol{X}_2^j \in \mathcal{X}^j\} + 0.5P\{g(t, \boldsymbol{X}_1^j) \\ &= g(t, \boldsymbol{X}_2^j) \mid G_1^{j+1} > t \geq G_2^{j+1}, \boldsymbol{X}_1^j \in \mathcal{X}^j, \boldsymbol{X}_2^j \in \mathcal{X}^j\},\end{aligned}$$

where $(\boldsymbol{X}_1^j, G_1^{j+1})$ and $(\boldsymbol{X}_2^j, G_2^{j+1})$ represent data from two independent subjects. A concordance of 1 indicates perfect prediction, while a value of 0.5 implies the score is no better than random guessing at ranking risk among individuals. A higher value of concordance is more desirable. Following McIntosh and Pepe (2002), it can be shown that $g(t, \boldsymbol{X}^j) = 1 - S^{j+1}(t \mid \boldsymbol{X}^j)$ yields the highest $\mathrm{CON}_t^{j+1}(g)$ among all possible functions g.

The performance of a risk score should be evaluated using a test or validation dataset that is independent of the dataset used to construct the score. For simplicity, we estimate the concordance using the observed sample, $\{\widetilde{G}_i^{j+1}, \Delta_i^{j+1}, \boldsymbol{X}_i^j\delta_i^j, \delta_i^j, C_i, \boldsymbol{W}_i; i = 1, \ldots, n\}$, without introducing additional notation for a new dataset:

$$\begin{aligned}&\widehat{\mathrm{CON}}_t^{j+1}(g) \\ &= \frac{\sum_{i,i'} v_{ii'}^{j+1} h(g(t, \boldsymbol{X}_i^j), g(t, \boldsymbol{X}_{i'}^j))I(\widetilde{G}_i^{j+1} > t \geq \widetilde{G}_{i'}^{j+1}, \boldsymbol{X}_i^j \in \mathcal{X}^j, \boldsymbol{X}_{i'}^j \in \mathcal{X}^j)}{\sum_{i,i'} v_{ii'}^{j+1} I(\widetilde{G}_i^{j+1} > t \geq \widetilde{G}_{i'}^{j+1}, \boldsymbol{X}_i^j \in \mathcal{X}^j, \boldsymbol{X}_{i'}^j \in \mathcal{X}^j)},\end{aligned} \quad (6)$$

where $h(a, b) = I(a < b) + 0.5I(a = b)$ and $v_{ii'}^{j+1} = \Delta_{i'}^{j+1}\widehat{S}_C(Y_i^j + t \mid \boldsymbol{W}_i)^{-1}\widehat{S}_C(Y_{i'}^j + \widetilde{G}_{i'}^{j+1} \mid \boldsymbol{W}_{i'})^{-1}$. Here we count concordant pairs in the numerator and the total number of comparable pairs in the denominator, incorporating IPCW to address censoring. For an overall evaluation of prediction accuracy, we calculate the concordance at a series of time points $t_1, t_2, \ldots, t_N$ and use the weighted average $\sum_{k=1}^N \omega(t_k)\widehat{\mathrm{CON}}_{t_k}^{j+1}(g)/\sum_{k=1}^N \omega(t_k)$, where $\omega(\cdot)$ is a prespecified weight function and allows for emphasizing specific time points.

The concordance metric can also be used as the splitting criterion to guide the tree-building process. We define the goodness of a split as the increase in concordance of the partition-based risk score in a parent node τ when it is split into child nodes τ_L and τ_R. The local concordance in τ is defined as $\mathrm{CON}_t^{j+1}(g; \tau) = P\{g(t, \boldsymbol{X}_1^j) < g(t, \boldsymbol{X}_2^j) \mid G_1^{j+1} > t \geq G_2^{j+1}, \boldsymbol{X}_1^j \in \tau, \boldsymbol{X}_2^j \in \tau\} + 0.5P\{g(t, \boldsymbol{X}_1^j) = g(t, \boldsymbol{X}_2^j) \mid G_1^{j+1} > t \geq G_2^{j+1}, \boldsymbol{X}_1^j \in \tau, \boldsymbol{X}_2^j \in \tau\}$. Instead of employing (6), we simplify the calculation by using the fact that the partition-based risk score is constant within a node. For any $\boldsymbol{x}^j$ within a node, $g(t, \boldsymbol{x}^j)$ is assigned the conditional probability of $G^{j+1} \leq t$ given $\boldsymbol{X}^j$ belongs to that node. Before splitting τ, $\mathrm{CON}_t(g; \tau)$ is 0.5. After splitting, $g(t, \boldsymbol{x}^j)$ is updated to reflect probabilities conditioned on membership in τ_L or τ_R. This leads to an increase in local concordance, expressed as

$$\frac{\begin{aligned}&|P(G^{j+1} > t, \boldsymbol{X}^j \in \tau_L)P(G^{j+1} \leq t, \boldsymbol{X}^j \in \tau_R) \\ &-P(G^{j+1} > t, \boldsymbol{X}^j \in \tau_R)P(G^{j+1} \leq t, \boldsymbol{X}^j \in \tau_L)|\end{aligned}}{P(G^{j+1} > t, \boldsymbol{X}^j \in \tau)P(G^{j+1} \leq t, \boldsymbol{X}^j \in \tau)}.$$

We estimate the probabilities involved in this formula using the observed data:

$$\begin{aligned}\widehat{P}(G^{j+1} > t, \boldsymbol{X}^j \in \tau) &= \frac{1}{n}\sum_{i=1}^{n} I(\widetilde{G}_i^{j+1} > t, \boldsymbol{X}_i^j \in \tau)/\widehat{S}_C(Y_i^j + t \mid \boldsymbol{W}_i), \\ &= \frac{1}{n}\sum_{i=1}^{n} \delta_i^j I(\boldsymbol{X}_i^j \in \tau)/\widehat{S}_C(Y_i^j \mid \boldsymbol{W}_i) - \widehat{P}(G^{j+1} > t, \boldsymbol{X}^j \in \tau).\end{aligned}$$

Lastly, to obtain an overall evaluation of the goodness-of-split, we compute the increase in concordance at a series of time points and use an average of the increase.

4. Global Prediction

The predictor space of enhanced global model, when incorporating LoS, is $\{\boldsymbol{z}^l = (\boldsymbol{w}^\top, t^l, g^l, u^l, \ldots, g^{l-m+1}, u^{l-m+1}, l)^\top : \boldsymbol{w} \in \mathcal{W}, t^l \in [0, t_{\max}], g^{l-k+1} \in [0, t_{\max}] \cup \{\text{NA}\}, u^{l-k+1} \in [0, t_{\max}] \cup \{\text{NA}\}, k = 1, \ldots, m, \sum_{k=1}^{\min(l,m)} (g^{l-k+1} + u^{l-k+1}) \leq t^l, l = 0, \ldots, q-1\}$. In decision trees, the splitting rule can readily handle NA values in predictors. To split a node based on a predictor Z that can take the NA value, we consider two candidate splits for each cutoff value c: (a) $\{Z > c\}$ versus $\{Z \leq c$ or $Z = \text{NA}\}$ and (b) $\{Z > c$ or $Z = \text{NA}\}$ versus $\{Z \leq c\}$. The split resulting in the largest improvement in the splitting criterion is chosen. The predictor space of the partially global model, $\{\boldsymbol{z}^l = (\boldsymbol{w}^\top, t^l, g^l, u^l, \ldots, g^{l-m+1}, u^{l-m+1}, l)^\top : \boldsymbol{w} \in \mathcal{W}, 0 \leq t^l, g^{l-k+1}, u^{l-k+1} \leq t_{\max}, k = 1, \ldots, m, \sum_{k=1}^{m}(g^{l-k+1} + u^{l-k+1}) \leq t^l, l = m, \ldots, q-1\}$, does not involve NA, so splitting follows conventional decision trees.

In the presence of censoring, we represent the training data in the long format as $\{\widetilde{G}_i^{j+1}, \Delta_i^{j+1}, \boldsymbol{Z}_i^j \delta_i^j, \boldsymbol{W}_i, C_i; i = 1, \ldots, n, j = 0, \ldots, q-1\}$, where $\boldsymbol{Z}^j = (\boldsymbol{W}^\top, T^j, H(j,m)^\top, j)^\top$. Estimation of partially and enhanced global models follows the same general approach. For ease of discussion, we denote the support of J as $\mathcal{J}$ and the predictor space as $\mathcal{Z}$ for both models. We denote the conditional hazard function of G^{J+1} given $\boldsymbol{Z}^J = \boldsymbol{z}^l$ as $\lambda(t \mid \boldsymbol{z}^l)dt = P\{G^{J+1} \in [t, t+dt) \mid G^{J+1} \geq t, \boldsymbol{Z}^J = \boldsymbol{z}^l\}$. Then we have $S(t \mid \boldsymbol{z}^l) = \exp\left\{-\int_0^t \lambda(u \mid \boldsymbol{z}^l)du\right\}$. For a node $\tau \subseteq \mathcal{Z}$, we define the corresponding within-node hazard function as

$$\lambda(t \mid \tau)dt = P\{G^{J+1} \in [t, t+dt) \mid G^{J+1} \geq t, \boldsymbol{Z}^J \in \tau\},$$

where J follows a discrete uniform distribution over $\mathcal{J}$. While our definition of $S(t \mid \boldsymbol{z}^l)$ conditions on $J = l$ and does not require specifying the distribution of J, the joint distribution of $\{G^{J+1}, \boldsymbol{Z}^J\}$ in τ depends on the distribution of J when τ contains multiple values of J. Therefore, we explicitly specify the distribution of J in the definition of $\lambda(t \mid \tau)$.

The recursive partitioning procedure in Section 3 is designed for independent observations in the training data. To account for possibly correlated observations while applying algorithms for independent observations, we consider within-cluster resampling (Hoffman, Sen, and Weinberg 2001; Follmann, Proschan, and Leifer 2003). This approach generates resampled datasets of independent observations by randomly selecting one observation from each subject. For the bth resampled dataset, $b = 1, \ldots, B$, suppose that the J_{bi}th observation of the ith subject is sampled. Given the random selection process, J_{bi} follows a discrete uniform distribution over $\mathcal{J}$ and is independent of other variables. The bth resampled dataset is represented as $\{\widetilde{G}_i^{J_{bi}+1}, \Delta_i^{J_{bi}+1}, \boldsymbol{Z}_i^{J_{bi}} \delta_i^{J_{bi}}, \boldsymbol{W}_i, C_i; i = 1, \ldots, n\}$, where $\{\widetilde{G}_i^{J_{bi}+1}, \Delta_i^{J_{bi}+1}, \boldsymbol{Z}_i^{J_{bi}} \delta_i^{J_{bi}}\}$ takes the value of $\{\widetilde{G}_i^{j+1}, \Delta_i^{j+1}, \boldsymbol{Z}_i^j \delta_i^j\}$ when $J_{bi} = j$. Using this dataset, we construct the partition function $\ell_b(\boldsymbol{z}^l)$, which identifies the terminal node containing $\boldsymbol{z}^l$ in the bth tree. The splitting criterion, which is based on concordance, is detailed in the supplementary materials.

In the absence of censoring, observations in the bth resampled dataset, $\{(G_i^{J_{bi}+1}, \boldsymbol{Z}_i^{J_{bi}}), i = 1, 2, \ldots, n\}$, share the same joint distribution as $(G^{J+1}, \boldsymbol{Z}^J)$. Then we have

$$\begin{aligned}&E\{I(\boldsymbol{Z}_i^{J_{bi}} \in \tau)dI(G_i^{J_{bi}+1} \leq t)\} \\ &\quad - E\{I(\boldsymbol{Z}_i^{J_{bi}} \in \tau)I(G_i^{J_{bi}+1} \geq t)\}\lambda(t \mid \tau)dt = 0.\end{aligned}$$

Define an indicator variable w_{bij} such that $w_{bij} = 1$ when the jth observation of the ith subject is included in the bth resampled dataset (i.e., $J_{bi} = j$), and $w_{bij} = 0$ otherwise. Under within cluster resampling, w_{bij} is independent of the observed data and $E(w_{bij}) = 1/|\mathcal{J}|$, where $|\mathcal{J}|$ is the number of elements in $\mathcal{J}$. An unbiased estimating equation for $\lambda(t \mid \tau)$ based on the bth resampled dataset is given by: $\sum_{i=1}^{n} I(\boldsymbol{Z}_i^{J_{bi}} \in \tau)\{dI(G_i^{J_{bi}+1} \leq t) - I(G_i^{J_{bi}+1} \geq t)\lambda(t \mid \tau)dt\} = 0$, or equivalently,

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j\in\mathcal{J}} w_{bij} I(\boldsymbol{Z}_i^j \in \tau)\{dI(G_i^{j+1} \leq t) \\ &\quad - I(G_i^{j+1} \geq t)\lambda(t \mid \tau)dt\} = 0. \qquad (7)\end{aligned}$$

To derive the ensemble estimating equation for $\lambda(t \mid \boldsymbol{z}^l)dt$ with partitions ℓ_b, $b = 1, 2, \ldots, B$, we replace τ with $\ell_b(\boldsymbol{z}^l)$ and average over estimating equations from all resampled datasets:

$$\begin{aligned}&\sum_{b=1}^{B}\sum_{i=1}^{n}\sum_{j\in\mathcal{J}} w_{bij} I(\boldsymbol{Z}_i^j \in \ell_b(\boldsymbol{z}^l))\{dI(G_i^{j+1} \leq t) \\ &\quad - I(G_i^{j+1} \geq t)\lambda(t \mid \boldsymbol{z}^l)dt\} = 0. \qquad (8)\end{aligned}$$

In the presence of censoring, we apply the IPCW technique, similar to the method in Section 3. Specifically, we have $E\{w_{bij} I(\boldsymbol{Z}_i^j \in \ell_b(\boldsymbol{z}^l))dN^{j+1}(t)/S_C(Y_i^j + t \mid \boldsymbol{W}_i)\} = E\{w_{bij} I(\boldsymbol{Z}_i^j \in \ell_b(\boldsymbol{z}^l))dI(G^{j+1} \leq t)\}$ and $E\{w_{bij} I(\boldsymbol{Z}_i^j \in \ell_b(\boldsymbol{z}^l))I(\widetilde{G}^{j+1} \geq t)/S_C(Y_i^j + t \mid \boldsymbol{W}_i)\} = E\{w_{bij} I(\boldsymbol{Z}_i^j \in \ell_b(\boldsymbol{z}^l))I(G^{j+1} \geq t)\}$. Based on (8), we construct the following estimating equation:

$$\begin{aligned}&\sum_{b=1}^{B}\sum_{i=1}^{n}\sum_{j\in\mathcal{J}} w_{bij} I(\boldsymbol{Z}_i^j \in \ell_b(\boldsymbol{z}^l))/\widehat{S}_C(Y_i^j + t \mid \boldsymbol{W}_i)\{dN_i^{j+1}(t) \\ &\quad - I(\widetilde{G}_i^{j+1} \geq t)\lambda(t \mid \boldsymbol{z}^l)dt\} = 0.\end{aligned}$$

Solving the estimating equation yields our estimator for the target survival function:

$$\widehat{S}(t \mid \boldsymbol{z}^l) = \exp\left\{-\int_0^t \frac{\sum_{i=1}^{n}\sum_{j\in\mathcal{J}} w_{ij}(\boldsymbol{z}^l, u)dN_i^{j+1}(u)}{\sum_{i=1}^{n}\sum_{j\in\mathcal{J}} w_{ij}(\boldsymbol{z}^l, u)I(\widetilde{G}_i^{j+1} \geq u)}\right\}, \qquad (9)$$

where $w_{ij}(\boldsymbol{z}^l, u) = B^{-1}\sum_{b=1}^{B} w_{bij}\delta_i^j I(\boldsymbol{Z}_i^j \in \ell_b(\boldsymbol{z}^l))/\widehat{S}_C(Y_i^j + u \mid \boldsymbol{W}_i)$. As indicated in (9), data from multiple episodes can be leveraged to estimate the risk of each event.

As an alternative to within-cluster resampling, one can consider subsampling all $n|\mathcal{J}|$ observations without replacement. In subsampling, we continue to use w_{bij} to indicate the inclusion of the jth observation of the ith subject in the bth subsampled dataset. Under this sampling scheme, w_{bij} is independent of the data and has a constant mean that does not depend on b, i, or j. In the supplementary materials, we prove that the estimating equation in (7) remains valid under subsampling. This allows us to derive the estimator $\widehat{S}(t \mid \boldsymbol{z}^l)$ in (9), following the same arguments as described above. Furthermore, the splitting criterion, which is the key component of the tree-building algorithm, can be represented as a continuous functional (with respect to the supremum norm) of the means of specific stochastic processes. Although the subsampled datasets contain correlated observations, we demonstrate in the supplementary materials that the estimates for these means are consistent, and the estimation of splitting criterion remains valid. This ensures that our method is reliable under subsampling. The subsampling approach can lead to larger resampled datasets compared to within-cluster resampling, enabling the construction of larger trees capable of handling more complex true models. The implementation of global models with different sampling strategies is outlined in Algorithm 1.

Algorithm 1 Global prediction

1: **Input:** Training data $\mathcal{D} = \{\widetilde{G}_i^{j+1}, \Delta_i^{j+1}, \boldsymbol{Z}_i^j\delta_i^j, \boldsymbol{W}_i, C_i; j \in \mathcal{J}, i = 1, \ldots, n\}$

- For the partially global model, set $\mathcal{J} = \{m, m+1, \ldots, q-1\}$ for a given $m > 0$
- For the enhanced global model, set $\mathcal{J} = \{0, 1, \ldots, q-1\}$

2: **Output:** Survival probability $\widehat{S}(t \mid \boldsymbol{z}^l)$ for a new observation with $\boldsymbol{Z}^l = \boldsymbol{z}^l$

3: **for** $b = 1, \ldots, B$ **do**

4: Generate the bth resampled dataset

- Within cluster resampling: randomly select one observation from each subject
- Subsampling: randomly select a subset of observations from $\mathcal{D}$ without replacement

5: Set $w_{bij} = 1$ if the jth observation of the ith individual is selected ($i = 1, \ldots, n, j \in \mathcal{J}$)

6: Construct ℓ_b from the resampled dataset via concordance-based splitting; at each split, select a random subset of *mtry* predictors as candidate splitting variables

7: **end for**

8: Predict $S(t \mid \boldsymbol{z}^l)$ using

$$\widehat{S}(t \mid \boldsymbol{z}^l) = \exp\left\{-\sum_{i=1}^{n}\sum_{j\in\mathcal{J}} \frac{w_{ij}(\boldsymbol{z}^l, \widetilde{G}_i^{j+1})\Delta_i^{j+1}}{\sum_{i'=1}^{n}\sum_{j'\in\mathcal{J}} w_{i'j'}(\boldsymbol{z}^l, \widetilde{G}_i^{j+1}) I(\widetilde{G}_{i'}^{j'+1} \geq \widetilde{G}_i^{j+1})}\right\},$$

where $w_{ij}(\boldsymbol{z}^l, u) = B^{-1}\sum_{b=1}^{B} w_{bij}\delta_i^j I(\boldsymbol{Z}_i^j \in \ell_b(\boldsymbol{z}^l))/\widehat{S}_C(Y_i^j + u \mid \boldsymbol{W}_i)$.

5. Simulation

We evaluated the performance of the proposed methods through simulations. The baseline predictors are $\boldsymbol{W} = (W_1, \ldots, W_{20})^\top$, where W_i followed the standard normal distribution for $i = 1, \ldots, 10$, and the Bernoulli distribution with a success probability of 0.5 for $i = 11, \ldots, 20$. The correlations between the baseline predictors were introduced using a copula approach (detailed in the supplementary materials) and were structured to decrease as the distance between variable indices increased. Aligned with the SEER-Medicare data, our main simulations focused on nonzero LoS. Simulations for recurrent event data with zero LoS are included in the supplementary materials. We set $q = 5$ and examined the prediction of $G^2, \ldots, G^5$. Gap times were simulated using the following models:

(M1) Gap times depended on different baseline predictors and the average time between previous hospital discharges: $\log(G^{j+1}) = 6 + W_{j+1}^2 - \sqrt{|W_{j+6}|} + W_{j+11} + (1.5 - j)\log(1 + T^j/j)/20 + \epsilon^{j+1}$ for $1 \leq j \leq 4$, and $\log(G^1) = 6 + W_1^2 - \sqrt{|W_6|} + W_{11} + \epsilon^1$.

(M2) Gap times depended on the same set of baseline predictors and the preceding hospitalization: $\log(G^{j+1}) = 6 + W_2^2 - \sqrt{|W_7|} + W_{12} + 0.25\log(1 + G^j) - \log(1 + U^j) + \epsilon^{j+1}$ for $1 \leq j \leq 4$, and $\log(G^1) = 6 + W_2^2 - \sqrt{|W_7|} + W_{12} + \epsilon^1$.

(M3) Gap times depended on the same set of baseline predictors and the first hospitalization: $\log(G^{j+1}) = 6 + W_2^2 - \sqrt{|W_7|} + W_{12} + 0.25\log(G^1) - \log(U^1) + \epsilon^{j+1}$, $1 \leq j \leq 4$, where $\log G^1$ followed a log-normal distribution with a mean of 6.5 and a standard deviation of 1.5 on the logarithmic scale.

The jth LoS, U^j, followed a gamma distribution with shape 2 and scale $0.7 - j/10$ in (M1) and (M2), and a log-normal distribution with mean 0 and standard deviation 0.5 on the logarithmic scale in (M3). The error terms ϵ^{j+1}, $0 \leq j \leq 4$ were independent and followed a standard normal distribution. In (M1) and (M2), the true global survival function $S(t \mid \boldsymbol{z}^l)$ does not depend on m for $m \geq 1$, and $S(t \mid \boldsymbol{Z}^j) = S^{j+1}(t \mid \boldsymbol{X}^j)$. In (M3), T^j was excluded from $\boldsymbol{Z}^j$ to ensure that the form of $S(t \mid \boldsymbol{z}^l)$ is computationally convenient. Moreover, $S(t \mid \boldsymbol{z}^l)$ in (M3) depends on m (see supplementary materials for details). For a given m, $S(t \mid \boldsymbol{Z}^j) = S^{j+1}(t \mid \boldsymbol{X}^j)$ holds only for $1 \leq j \leq m$ in the enhanced global model and $j = m$ in the partially global model, that is, when information on the first event is included in the global predictors. Plots of the distributions of gap times in (M1)–(M3) are shown in the supplementary materials. We independently generated C from an exponential distribution with a rate parameter that yielded an average of three observed events per subject. For each scenario, we performed a simulation run of 500 replicates.

We evaluated the prediction accuracy of three types of ensemble models: episode-specific, partially global, and enhanced global. Different values of m were explored in global

models. Each ensemble model consisted of 500 unpruned trees. At each split, $mtry = \lceil \sqrt{p} \rceil$ variables were randomly selected as candidate splitting variables, where p is the total number of predictors. For episode-specific models, trees were constructed using bootstrap samples. In global models, we employed within-cluster resampling and subsampling. The subsampling method selected 63.2% of the $n|\mathcal{J}|$ observations without replacement, where $|\mathcal{J}|$ denotes the cardinality of $\mathcal{J}$. The survival function of C was estimated using empirical probability. The minimum terminal node size was set to 5 for episode-specific models and global models employing within cluster resampling, and to 10 for global models with subsampling when $n = 100$. For $n = 200$, the minimum terminal node sizes were adjusted to 7 and 15, respectively. We set $t_{\max} = \infty$ in all models. As a comparison, we fitted episode-specific Cox-type models using subjects with $\delta^j = 1$. In these models, the hazard for G^{j+1} was modeled as $\lambda^{j+1}(t \mid \boldsymbol{X}^j) = \lambda_0^{j+1}(t)\exp(\boldsymbol{\beta}^{j\top}\boldsymbol{X}^j)$, where $\boldsymbol{\beta}^j$ is the vector of regression parameters and $\lambda_0^{j+1}(t)$ is an unspecified baseline hazard function.

In each simulation run, models were built using training data of sizes $n = 100$ and $n = 200$. We evaluated their performance using independent uncensored test data with 500 subjects, $\{\boldsymbol{W}_i^*, G_i^{j*}, U_i^{j*}, j = 1,\ldots,q, i = 1,\ldots,500\}$, where the superscript * indicates test data. For the prediction of G^{j+1}, mean squared error (MSE) and concordance (CON) at time t were defined as $\sum_{i=1}^{500}\{\widehat{S}^{j+1}(t \mid \boldsymbol{X}_i^{j*}) - S^{j+1}(t \mid \boldsymbol{X}_i^{j*})\}^2/500$ and $\widehat{\mathrm{CON}}_t^{j+1*}\{\widehat{S}^{j+1}(t \mid \boldsymbol{X}^{j*})\}$ in episode-specific prediction, and $\sum_{i=1}^{500}\{\widehat{S}(t \mid \boldsymbol{Z}_i^{j*}) - S(t \mid \boldsymbol{Z}_i^{j*})\}^2/500$ and $\widehat{\mathrm{CON}}_t^{j+1*}\{\widehat{S}(t \mid \boldsymbol{Z}^{j*})\}$ in global prediction, where $\widehat{\mathrm{CON}}_t^{j+1*}$ was estimated based on the test data. Prediction accuracy for G^{j+1} was evaluated using average MSE over 100 time points between 0 and $Q^{j+1}(0.9)$, and average CON over 100 time points between $Q^{j+1}(0.1)$ and $Q^{j+1}(0.9)$, where $Q^{j+1}(p)$ denotes the pth quantile of G^{j+1} among individuals with $\Delta^{j+1} = 1$. Cox-type models occasionally encountered convergence issues, particularly with $n = 100$ or when predicting later gap times. These nonconvergence instances were excluded from the MSE and CON evaluations.

The average MSE and average CON are presented in Table 1 and Table S1 in the supplementary materials, respectively. The proposed models generally outperformed the Cox-type model, especially for predicting later gap times. In (M1), where limited information can be borrowed across episodes, the episode-specific model performed better than or similar to the global models. In contrast, in (M2), the global models effectively leveraged information across episodes and outperformed the episode-specific model. For (M3), average MSE comparisons with episode-specific models are meaningful for $G^2,\ldots,G^{m+1}$ in enhanced global models and for G^{m+1} in partially global models, while average CON can be used to compare all models for predicting any gap times. Among valid comparisons, the enhanced global model generally outperformed the episode-specific model, while the partially global model performed better for G^2 and similarly for other events when compared to the episode-specific model. The enhanced global model with $m = 4$ achieved higher average CON than models with smaller m values, due to the inclusion of informative predictors G^1 and U^1 for all episodes. This pattern contrasts with (M1) and (M2), where no single m value consistently dominated others under the enhanced global model. In all scenarios, neither a larger nor a smaller m in partially global models consistently improved prediction accuracy, suggesting a potential tradeoff between using more historical information as predictors and using more observations for model training. For enhanced global models, subsampling consistently outperformed within-cluster resampling. In partially global models, the relative performance of subsampling versus within-cluster resampling varied with m. This difference in performance might arise from the size of the resampled datasets: subsampling in enhanced global models or partially global models with smaller m produced larger resampled datasets, allowing for the growth of larger trees that better capture the complexity of the true model. Additional simulations that explored increasing correlations among episodes, varying distributions of the number of observed events, and transformed predictors are provided in the supplementary materials.

6. Data Application

BC creates a significant health burden for women globally. While advancements in diagnosis and treatment have resulted in improved survival rates for BC patients, management of comorbid conditions continues to pose significant challenges, particularly among elderly patients. Using SEER-Medicare data, we developed prediction models for recurrent hospitalizations in elderly BC patients. These models aim to enable healthcare providers to implement timely interventions and tailor healthcare strategies to improve patient outcomes. Our analysis focused on female Medicare beneficiaries aged 66 or older diagnosed with early-stage BC from 2010 to 2017. We designated the date of BC diagnosis as the baseline (time zero). The follow-up ended on December 1, 2018, or at death, whichever came first. To ensure recurrent hospitalizations are recorded, our inclusion criteria further required continuous enrollment in Medicare Parts A and B for at least 12 months before and throughout the follow-up period. According to these criteria, we randomly selected 3000 patients for training data and another 3000 for test data. We set $q = 5$ to predict the first five recurrent hospitalizations after BC diagnosis. Hospital admission was defined as the event of interest, and the prediction focused on the gap time between discharge from one hospital stay and subsequent admission.

Our models included two types of predictors: baseline patient characteristics and time-varying historical information. Baseline predictors included sociodemographic factors such as age at baseline, race (White, Black or African American, and other), ethnicity (Hispanic and non-Hispanic), marital status (never married, married, and other), and clinical factors such as tumor size in millimeters, cancer stage (I, II, and III/IV), BC subtype (HER2+/HR+, HER2+/HR−, HER2-/HR+, and triple-negative), and lymph node examination results (all negative, no nodes examined, and some positive). Historical information included both previous gap times and LoS, with LoS serving as a proxy for illness severity. Descriptive statistics of predictors and gap times are presented in supplementary materials.

For comparison, we included a set of episode-specific Cox-type models. Figure S1 in the supplementary materials

Table 1. Average MSE ($\times 1000$) for simulation scenarios M1, M2, and M3.

	Scenario M1								Scenario M2								Scenario M3							
	G^2	G^3	G^4	G^5	G^2	G^3	G^4	G^5	G^2	G^3	G^4	G^5	G^2	G^3	G^4	G^5	G^2	G^3	G^4	G^5	G^2	G^3	G^4	G^5
	$n = 100$				$n = 200$				$n = 100$				$n = 200$				$n = 100$				$n = 200$			
Cox model																								
	134	167	199	255	103	121	136	161	69	78	105	154	46	46	52	69	124	93	107	152	97	56	56	67
Episode-specific model																								
	84	86	86	93	70	75	74	85	67	68	71	75	54	54	57	60	95	66	61	59	82	54	48	46
Partially global model																								
Subsampling																								
$m = 1$	79	84	88	101	76	80	83	97	53	53	54	57	44	44	45	48	59	50	47	45	52	40	37	36
$m = 2$		78	82	93		75	79	91		58	55	56		47	46	47		65	44	43		56	35	34
$m = 3$			78	86			73	82			68	65			57	55			61	46			51	35
Within cluster resampling																								
$m = 1$	80	85	91	105	77	82	87	101	58	57	58	61	47	47	48	51	62	53	50	49	55	44	40	39
$m = 2$		79	84	95		76	80	92		59	56	57		47	46	47		66	45	44		57	35	34
$m = 3$			78	87			73	82			63	60			51	49			58	42			47	31
Enhanced global model																								
Subsampling																								
$m = 1$	80	86	90	102	77	84	86	98	47	47	46	46	38	38	37	38	61	47	45	43	56	38	36	34
$m = 2$	82	82	87	99	79	79	83	95	51	45	44	45	42	36	36	37	65	58	39	37	59	50	29	28
$m = 3$	82	86	83	94	79	84	79	90	52	47	44	45	43	38	36	37	65	60	51	38	60	52	42	29
$m = 4$	82	87	84	91	79	83	80	86	52	46	44	44	43	38	36	36	65	60	51	49	60	51	41	39
Within cluster resampling																								
$m = 1$	81	87	94	108	79	85	90	104	55	54	53	55	44	43	42	44	64	53	50	48	59	43	40	39
$m = 2$	85	83	90	104	84	81	87	100	61	53	52	53	49	42	41	42	67	65	45	44	63	55	34	33
$m = 3$	85	88	86	98	84	87	82	94	62	55	53	53	51	44	41	42	68	67	58	45	63	57	47	34
$m = 4$	85	88	87	97	84	87	84	92	62	55	52	53	50	44	41	42	68	67	58	56	63	57	47	45

summarizes the results from these models for the second to fifth gap times. As fewer participants remained at risk for later events, the confidence intervals for the hazard ratios tended to widen. Black patients (vs. White patients) and patients with no lymph nodes examined (vs. all-negative nodes) had an increased risk for the second event. However, these predictors were not significantly associated with later events. The relationship between event risks and hospitalization history was complex. For instance, a longer first gap time increased the risk for the second event but reduced the risk for the fourth event, and a longer first LoS increased the risk for the fourth event. Despite their interpretability, Cox-type models offered limited insights into these complex patterns. Some predictors, such as race (Black) and the first LoS, consistently had hazard ratios greater than one, suggesting similar effects across episodes.

Given multiple modeling strategies and various choices of m, we applied 10-fold cross-validation to select the optimal model. Prediction performance was assessed using concordance. For each gap time (G^2 through G^5) and model, we calculated the average time-dependent concordance over 100 evenly spaced time points within the 10th to 90th percentile range of observed uncensored gap times. The overall concordance for each model was then calculated as the average of the episode-specific concordance values for the four gap times. For the partially global model that predicts the $(m+1)$th and subsequent gap times, the average concordance for G^{j+1} with $1 \leq j < m$ was calculated using the respective episode-specific models. Table 2 presents the cross-validation results and test data performance for all models. The partially global model with $m = 1$ and subsampling achieved the highest overall concordance in cross-validation and was selected as the final model. It also achieved the highest overall concordance on the test data. Compared to tree-based methods, the Cox-type models showed limited predictive accuracy across all gap times. The episode-specific models had lower concordance for later gap times compared to global models, suggesting that borrowing information from earlier episodes may be beneficial.

Due to the challenge of visualizing functions with many arguments, we explored the partial dependence of $\widehat{S}(t \mid \boldsymbol{z}^l)$ on a small subset of predictors (Friedman 2001). Here, we fix $J = l$ and consider an episode-specific definition of partial dependence for $\widehat{S}(t \mid \boldsymbol{z}^l)$. Let $\boldsymbol{Z}^l$ denote the predictors for the $(l+1)$th event in a global model. We divide $\boldsymbol{Z}^l$ into a selected subset $\boldsymbol{Z}_s^l$ and a complement subset $\boldsymbol{Z}_c^l$ (the episode number l is included in $\boldsymbol{Z}_c^l$). Without loss of generality, we express $\boldsymbol{Z}^l$ as $(\boldsymbol{Z}_s^{l\top}, \boldsymbol{Z}_c^{l\top})^\top$. For the $(l+1)$th event, the partial dependence function of $\boldsymbol{Z}_s^l$ is defined as $\bar{S}^{l+1}(t, \boldsymbol{z}_s^l) = E_{\boldsymbol{Z}_c^l}\{\widehat{S}(t \mid (\boldsymbol{z}_s^{l\top}, \boldsymbol{Z}_c^{l\top})^\top) I(T^l \leq t_{\max})\}$, which captures how the risk of the $(l+1)$th event depends on $\boldsymbol{Z}_s^l$ after integrating over the distribution of $\boldsymbol{Z}_c^l$. In the presence of censoring, we estimate $\bar{S}^{l+1}(t, \boldsymbol{z}_s^l)$ using $n^{-1}\sum_{i=1}^n \delta_i^l \widehat{S}_C(Y_i^l \mid \boldsymbol{W}_i)^{-1} \widehat{S}(t \mid (\boldsymbol{z}_s^{l\top}, \boldsymbol{Z}_{c,i}^{l\top})^\top) I(Y_i^l \leq t_{\max})$.

We generated partial dependence plots with $\boldsymbol{Z}_s^l$ selected as two continuous predictors: LoS and gap time of the most recent hospitalization. We employed $\widehat{S}(t \mid \boldsymbol{z}^l)$ from the partially global model with subsampling and $m = 1$. Figure 4 presents the partial dependence function evaluated at $t = 100$ days for the second to fifth hospitalizations ($l = 1, \ldots, 4$). The magnitude of

Table 2. Cross-validation and test data concordance for the SEER-Medicare data.

	10-fold cross-validation										Test data									
	G^2	G^3	G^4	G^5	Overall	G^2	G^3	G^4	G^5	Overall	G^2	G^3	G^4	G^5	Overall	G^2	G^3	G^4	G^5	Overall
	Cox-type					Episode-specific					Cox-type					Episode-specific				
	601	588	562	636	(597)	636	631	598	658	(631)	584	591	542	660	(594)	608	623	580	686	(624)
		Partially global					Enhanced global					Partially global					Enhanced global			
	Subsampling																			
$m=1$	644	649	675	727	(674)	630	601	636	711	(645)	620	637	649	729	(659)	620	594	626	714	(639)
$m=2$		607	582	633	(614)	624	587	651	740	(651)		598	548	642	(599)	614	591	650	741	(649)
$m=3$			583	720	(643)	617	581	670	748	(654)			564	737	(633)	608	583	666	749	(652)
$m=4$						615	578	666	724	(646)						607	581	664	729	(645)
	Within cluster resampling										Within cluster resampling									
$m=1$	645	646	665	707	(666)	605	599	686	692	(646)	622	634	639	713	(627)	579	591	671	688	(633)
$m=2$		623	623	630	(628)	603	573	666	722	(641)		622	613	637	(620)	583	562	660	727	(633)
$m=3$			603	697	(642)	593	583	658	759	(648)			591	721	(636)	574	579	663	758	(644)
$m=4$						596	596	680	708	(645)						581	596	682	721	(645)

NOTE: The values are scaled by a factor of 1000. The overall concordance for each model, averaged across the four gap times, is shown in parentheses. For the overall concordance of partially global models, the concordances for G^2 to G^m were calculated based on episode-specific models.

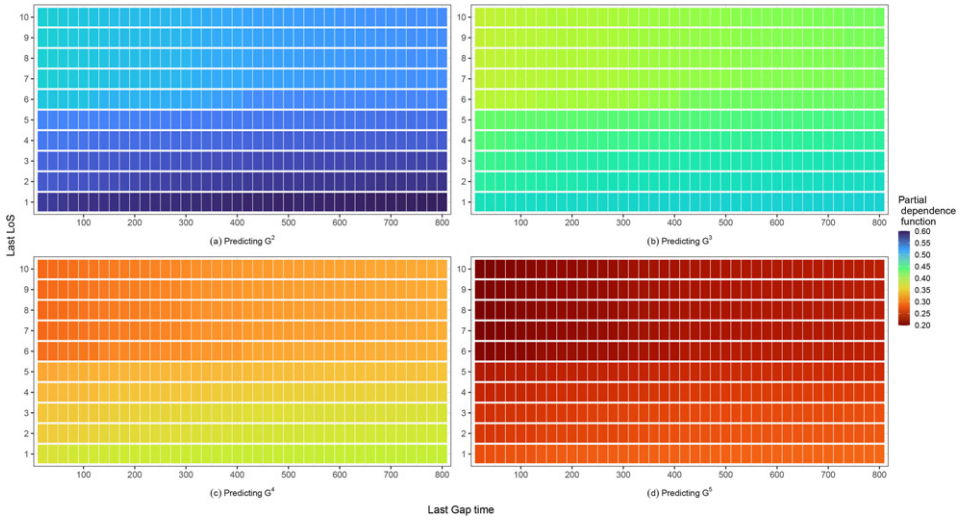


Figure 4. Partial dependence plots of event-free probability at 100 days with varying LoS and gap time of the most recent hospitalization.

risk varied across events, generally increasing with later hospitalizations. On the other hand, the plots show a similar pattern across these hospitalizations: across all plots, a cutoff point is observed around 6 days for LoS, with a noticeable decrease in risk for patients whose LoS was less than 6 days. Furthermore, for G^2, the impact of gap time became more pronounced when the LoS exceeded 6 days. To explore potential interactions between LoS and gap time, we conducted a two-way ANOVA on the logit-transformed predicted probabilities. We considered two factors: whether the preceding gap time was less than 90 days and whether the preceding LoS was within one week. The interaction terms for the second to fifth gap times yielded p-values of 0.034, 0.186, 0.033, and 0.882, respectively, suggesting potential interactions for the second and fourth gap times. Additional plots illustrating the interaction effects are presented in the supplementary materials.

7. Discussion

In this article, we develop a unified framework for the dynamic prediction of recurrent events and propose two modeling strategies: episode-specific prediction and global prediction. The global prediction strategy borrows information across events and is expected to improve prediction accuracy when multiple events share similarities in their relationship to predictors. The effectiveness of each strategy depends on the true underlying model and should be assessed in the context of each application. In addition to data-driven approaches such as cross-validation, domain knowledge can be valuable in guiding model selection. For instance, in a population primarily hospitalized due to chronic conditions like cardiovascular disease, where repeated admissions share consistent risk factors such as age and comorbidities, a global model may be appropriate. In contrast, episode-specific models may be more suitable when the factors that influence each event vary over time. For example, patients with multiple conditions or undergoing progressive treatments can experience changes in risk factors between hospitalizations. In these cases, episode-specific models allow the consideration of unique risk factors specific to each event.

In the absence of censoring, our method can be readily applied to dynamically predict longitudinal outcomes. In this case, G^{j+1} represents the outcome variable, such as a biomarker measurement. We can then incorporate the historical trajectory as predictors. Moreover, instead of the survival function, the mean can be predicted.

Supplementary Materials

The supplementary materials include additional methodological details on the episode-specific and global models, and further simulation and data analysis results. R code is available at: *https://github.com/stc04003/reForest*.

Acknowledgments

The collection of cancer incidence data used in this study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885; Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries, under cooperative agreement 1NU58DP007156; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201800032I awarded to the University of California, San Francisco, contract HHSN2612018000015I awarded to the University of Southern California, and contract HHSN2612018000009I awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the author(s) and do not necessarily reflect the opinions of the State of California, the Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

Yifei Sun gratefully acknowledges support from the National Institutes of Health (grants R21HL156228 and RF1AG081413).

References

Andersen, P. K., and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10, 1100–1120. [2015]

Beam, A. L., and Kohane, I. S. (2018), "Big Data and Machine Learning in Health Care," *Journal of the American Medical Association*, 319, 1317–1318. [2015]

Breiman, L. (2001), "Random Forests," *Machine Learning*, 45, 5–32. [2020]
Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984), *Classification and Regression Trees*, New York: Chapman & Hall. [2016,2019]
Chang, I.-S., and Hsiung, C. A. (1994), "Information and Asymptotic Efficiency in Some Generalized Proportional Hazards Models for Counting Processes," *The Annals of Statistics*, 22, 1275–1298. [2016]
Chang, S.-H., and Wang, M.-C. (1999), "Conditional Regression Analysis for Recurrence Time Data," *Journal of the American Statistical Association*, 94, 1221–1230. [2016]
Davis, R. B., and Anderson, J. R. (1989), "Exponential Survival Trees," *Statistics in Medicine*, 8, 947–961. [2016]
Enewold, L., Parsons, H., Zhao, L., Bott, D., Rivera, D. R., Barrett, M. J., Virnig, B. A., and Warren, J. L. (2020), "Updated Overview of the SEER-Medicare Data: Enhanced Content and Applications," *Jounral of the National Cancer Institute Monographs*, 2020, 3–13. [2015]
Follmann, D., Proschan, M., and Leifer, E. (2003), "Multiple Outputation: Inference for Complex Clustered Data by Averaging Analyses from Independent Data," *Biometrics*, 59, 420–429. [2016,2021]
Friedman, J. H. (2001), "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 29, 1189–1232. [2025]
Gordon, L., and Olshen, R. A. (1985), "Tree-Structured Survival Analysis," *Cancer Treatment Reports*, 69, 1065–1069. [2016]
Hoffman, E. B., Sen, P. K., and Weinberg, C. R. (2001), "Within-Cluster Resampling," *Biometrika*, 88, 1121–1134. [2016,2021]
Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., and Van Der Laan, M. J. (2006), "Survival Ensembles," *Biostatistics*, 7, 355–373. [2016]
Hothorn, T., Lausen, B., Benner, A., and Radespiel-Tröger, M. (2004), "Bagging Survival Trees," *Statistics in Medicine*, 23, 77–91. [2016]
Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008), "Random Survival Forests," *The Annals of Applied Statistics*, 2, 841–860. [2016]
LeBlanc, M., and Crowley, J. (1992), "Relative Risk Trees for Censored Survival Data," *Biometrics*, 48, 411–425. [2016]
——— (1993), "Survival Trees by Goodness of Split," *Journal of the American Statistical Association*, 88, 457–467. [2016]
McIntosh, M. W., and Pepe, M. S. (2002), "Combining Several Screening Tests: Optimality of the Risk Score," *Biometrics*, 58, 657–664. [2020]
Molinaro, A. M., Dudoit, S., and Van der Laan, M. J. (2004), "Tree-based Multivariate Regression and Density Estimation with Right-Censored Data," *Journal of Multivariate Analysis*, 90, 154–177. [2016]
Nuggerud-Galeas, S., Sáez-Benito Suescun, L., Berenguer Torrijo, N., Sáez-Benito Suescun, A., Aguilar-Latorre, A., Magallón Botaya, R., and Oliván Blázquez, B. (2020), "Analysis of Depressive Episodes, their Recurrence and Pharmacologic Treatment in Primary Care Patients: A Retrospective Descriptive Study," *PloS One*, 15, e0233454. [2015]
Pedersen, R. N., Esen, B. Ö., Mellemkjær, L., Christiansen, P., Ejlertsen, B., Lash, T. L., Nørgaard, M., and Cronin-Fenton, D. (2022), "The Incidence of Breast Cancer Recurrence 10–32 Years After Primary Diagnosis," *Journal of the National Cancer Institute*, 114, 391–399. [2015]
Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981), "On the Regression Analysis of Multivariate Failure Time Data," *Biometrika*, 68, 373–379. [2015]
Segal, M. R. (1988), "Regression Trees for Censored Data," *Biometrics*, 44, 35–47. [2016]
Sparapani, R. A., Rein, L. E., Tarima, S. S., Jackson, T. A., and Meurer, J. R. (2020), "Non-Parametric Recurrent Events Analysis with BART and an Application to the Hospital Admissions of Patients with Diabetes," *Biostatistics*, 21, 69–85. [2016]
Steingrimsson, J. A., Diao, L., Molinaro, A. M., and Strawderman, R. L. (2016), "Doubly Robust Survival Trees," *Statistics in Medicine*, 35, 3595–3612. [2016]
Steingrimsson, J. A., Diao, L., and Strawderman, R. L. (2019), "Censoring Unbiased Regression Trees and Ensembles," *Journal of the American Statistical Association*, 114, 370–383. [2016]
Warren, J. L., Klabunde, C. N., Schrag, D., Bach, P. B., and Riley, G. F. (2002), "Overview of the SEER-Medicare Data: Content, Research Applications, and Generalizability to the United States Elderly Population," *Medical Care*, 40, IV3–IV18. [2015]
Zhu, R., and Kosorok, M. R. (2012), "Recursively Imputed Survival Trees," *Journal of the American Statistical Association*, 107, 331–340. [2016]