Summer 8-3-2022

Dynamic Prediction for Alternating Recurrent Events Using a Semiparametric Joint Frailty Model

Jaehyeon Yun
jaehyeony@smu.edu

DYNAMIC PREDICTION FOR ALTERNATING RECURRENT EVENTS USING A SEMIPARAMETRIC JOINT FRAILTY MODEL

Approved by:

Dr. Daniel F. Heitjan
Professor in Department of Statistical Science, SMU & Population and Data Sciences, UTSW

Dr. Jing Cao
Professor in Department of Statistical Science, SMU

Dr. Chul Moon
Assistant Professor in Department of Statistical Science, SMU

Dr. Yu-Lun Liu
Assistant Professor in Department of Population and Data Sciences, UTSW

DYNAMIC PREDICTION FOR ALTERNATING RECURRENT EVENTS USING A SEMIPARAMETRIC JOINT FRAILTY MODEL

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Statistics

by

Yun Jaehyeon

B.A., Economics, Sungkyunkwan University
M.S., Statistics, Sungkyunkwan University

August 3, 2022

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Heitjan. Choosing him as my advisor was the best choice I made during my PhD. During the last two and a half years, he not only taught me statistics, but also helped me to develop a positive and correct attitude towards learning. These will be the cornerstones of the rest of my life beyond PhD. Words cannot describe how grateful I am. Thank you to my parents and brother for sustaining me over the past five years. Without their unconditional love and understanding, this dissertation would not have been done. Dr. Cao, Dr. Moon, Dr. Liu and Dr. Ahuja, thank all of you for your time, encouragement, and comments to improve this dissertation. Special thanks to Sheila. You are the true hero of our department who has always stood behind and supported us. Lastly, I am deeply appreciative of my friends and all the people I have met at SMU. It has been a great journey. I hope that anyone who reads this dissertation enjoys a journey to the frontier of learning. A lingering question will be answered tomorrow.

Jaehyeon, Yun

B.A., Economics, Sungkyunkwan University
M.S., Statistics, Sungkyunkwan University

Dynamic Prediction for Alternating Recurrent Events
Using a Semiparametric Joint Frailty Model

Advisor: Dr. Daniel F. Heitjan
Doctor of Philosophy degree conferred August 3, 2022
Dissertation completed July 7, 2022

Alternating recurrent events data arise commonly in health research; examples include hospital admissions and discharges of diabetes patients; exacerbations and remissions of chronic bronchitis; and quitting and restarting smoking. Recent work has involved formulating and estimating joint models for the recurrent event times considering non-negligible event durations. However, prediction models for transition between recurrent events are lacking. We consider the development and evaluation of methods for predicting future events within these models. Specifically, we propose a tool for dynamically predicting transition between alternating recurrent events in real time. Under a flexible joint frailty model, we derive the predictive probability of a transition from one event type to the other within a pre-specified time period. To circumvent numerical integration in calculating the predictive probability, we obtain the approximate transition probability by a Taylor expansion. Simulation results demonstrate that our tool provides better prediction performance in discrimination, as measured by the area under the ROC curve (AUC) and sensitivity, than prediction approaches that rely on standard binary regression models. Also, simulation shows that prediction results from approximate transition probability are as close as results from the exact predictive probability. We illustrate predictions in analyses of relapses of chronic bronchitis exacerbation from a pharmaceutical trial and hospital readmissions in patients with diabetes from Medicaid claims data.

The final part of this dissertation (Chapter 6) compares predictive performance between logistic regression and random forests for 30-day readmission using longitudinal claims data. Several studies have compared these and other prediction models using longitudinal electronic health records or claims data. Because most of them applied logistic regression to the longitudinal observations, ignoring the lack of independence within subjects, or claims data consisting of independent observations, a correct comparison of the models under longitudinal data remains obscure. Moreover, those studies did not compare the out-of-sample performance. We address these issues and compare the prediction performance of the models using longitudinal claims data. We implement simulations by randomly choosing a record from each patient's multiple records in the training set, fitting the two models, applying the models to the training, test, and external sets, and obtaining AUC and sensitivity for each. We observe that although random forests generally gives better predictions on the training set, logistic regression performs better on test and external sets. In an empirical study, we apply the prediction methods to Medicaid claims data covering inpatient admissions of patients with heart failure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I dedicate this thesis to my family.

For their endless love, support, and encouragement.

CHAPTER 1

Introduction

Diabetes is a common and serious disease with a global prevalence of 9.3% [34] and a US prevalence of 13%. Hospitalized patients with diabetes are twice as likely to be re-admitted within 30 days of discharge compared to similar patients without diabetes [32]. Thus, these patients are highly likely to experience recurrences of admission and discharge alternatively and this makes it hard to maintain their daily life in a long term. If future readmission risk can be estimated at discharge, medical providers can create treatment plans that will forestall readmissions without compromising patient health. Also, this can improve patients prospects of remaining healthy at home and avoid repeated lengthy recurrences. In this dissertation up to Chapter 5, we construct a joint frailty model describing transition intensity of such alternating recurrent events and based on the model, propose a dynamic prediction tool for a future recurrence with a window of prediction.

If a patient has multiple hospitalization records, the alternating recurrent events for admission and discharge are nested in the patient with within-patient correlation between events. Other potential sources of correlation include medical providers, hospitals, and regions. Thus, modeling the transition intensity to account for cluster effects will presumably give rise to more accurate predictions.

Sud et al [37] found that longer hospital stays for heart failure are associated with increased risk of cardiovascular and heart failure readmission. Thus, we should include previous length of stay information in models for the readmission intensity. Naturally, constructing transition intensity models with alternating gap time sequences over total time is desirable [7].

This relationship can also manifest as a negative correlation between the lengths of stay and discharge in an alternating recurrent events model. Failure to account for this correlation can lead to biased model estimation and, *a fortiori*, inaccurate prediction.

Various statistical methods have been developed for alternating recurrent events. Xue and Brookmeyer [40] introduced a bivariate frailty model for the analysis of bivariate survival data. This model accommodates clustered multivariate failure time data with two possibly correlated frailties following a bivariate normal distribution. Instead of gap-time scale, this model focused on calendar-time scale for failure time. Cook et al. [6] developed a model for alternative recurrent events based on Weibull forms for the conditional transition intensities and bivariate normal frailties to accommodate subject variability. They adopted Weibull hazard functions for the baseline intensity to address trends of event duration in the transition intensities. Lawless et al. [18] considered the modeling and analysis of event durations, also called *sojourn times*, with semi-Markov models where the durations of different sojourns are assumed to be independent. Instead of a full parametric approach, they applied piecewise constant hazard functions for the baseline intensity to flexibly reflect the relationship between event durations and transition intensity. From these studies, Cook and Lawless [7] described a joint transition frailty model for alternating recurrent events. Their model uses a gap-time formulation and directly accounts for non-negligible event durations in the model via baseline intensity function. It moreover assumes a bivariate normal frailty distribution without restriction on the sign of the correlation. Later, Mazroui et al. [25] extended this to multivariate frailty models for two types of recurrent events with a dependent terminal event.

Various methods exist for estimating joint frailty models. Qing [19] proposed a pseudo-likelihood method that assumes a parametric model for the baseline intensity function. Wang et al. [38] applied penalized partial log-likelihood to estimate the regression coefficients and individual frailties. By avoiding the estimation of the baseline transition intensity, their method gives fast computation times even with large samples. Liu et al. [23], Huang et al. [14], and Rondeau et al. [31] proposed shared frailty models for recurrent events and a

terminal event. In order to estimate the parameters in two correlated intensity functions, they used an EM algorithm, Laplace approximation, or semiparametric penalized likelihood approach including cubic M-splines and a smoothing penalty for the baseline functions. Duchateau and Janssen [9] and Wen et al. [39] proposed Bayesian estimation using Markov chain Monte Carlo.

Although numerous models and estimation methods are now available, risk prediction for future recurrence of alternating recurrent events is still lacking. Some studies developed prediction tools using joint frailty models, but their interest is not a recurrent event (such as cancer relapse) but a terminal event (such as death) [27][30][24]. Li et al. [20] studied smoking cessation trials and proposed prediction tools for individual long-term smoking cessation success based on a cure-frailty model. To our knowledge, there has been no proposed prediction tool for alternating recurrent events under the joint frailty model.

The remainder of this dissertation is organized as follows: In Chapter 2, we define an alternating two-state process and joint transition intensity model for alternating recurrent events. We estimate the model parameters by adopting the penalized partial log-likelihood approach of Wang et al. [38]. Also, we propose nonparametric estimation of the baseline transition intensity function using the Breslow estimator and B-spline regression. Chapter 3 presents a dynamic prediction tool that considers process history. In Chapter 4, we use simulation to compare prediction performance between our approach and other methods that rely on standard binary regression models. In Chapter 5, we apply the proposed tool in two real-data applications: Exacerbation and remission of chronic bronchitis, and readmission of Medicaid patients with diabetes. Chapter 6 uses simulation and a real-data application to compare the in-sample and out-of-sample prediction performance of logistic regression and random forests. Chapter 7 gives some concluding remarks.

CHAPTER 2

Joint Survival Modeling for Alternating Recurrent Events

2.1. Alternating two-state process

An alternating two-state process is a special case of a multi-state process when there are only 2 possible states and neither is terminal. In Figure 2.1, there are two events indexed by 0 and 1 occurring alternatively; these are called alternating recurrent events.

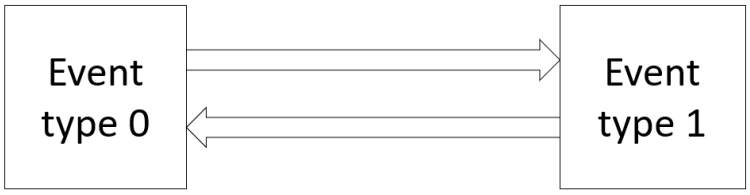

Figure 2.1: Alternating two-state process

Assume that there are n independent subjects being followed up over time who undergo alternating recurrent events. Let T^*_{ijk} denote the time of the kth occurrence of event type $j \in \{0, 1\}$ for subject $i \in \{1, \ldots, n\}$. We assume that every subject has at least one event for each type and the record always starts from type 0 and is censored with type 1. Consequently, the total event times are ordered by $0 < T^*_{i01} < T^*_{i11} < T^*_{i02} < T^*_{i12} < \cdots$. We define a gap time sequence between two alternating events. Given the total event times, $W^*_{i0k} = T^*_{i0k} - T^*_{i1(k-1)}$ is the gap time for $j = 0$, and $W^*_{i1k} = T^*_{i1k} - T^*_{i0k}$ is the gap time for $j = 1$. Note that when $k = 1$, T^*_{i10} is the starting point for recording, and so it is equal to 0.

The alternating two-state process is right-censored so that total observed event times and censoring indicator are defined as $T_{ijk} = \min(T^*_{ijk}, C_i)$ and $\delta_{ijk} = I(T^*_{ijk} \leq C_i)$ where C_i is a censoring time and $I(\cdot)$ is a 0/1 indicator function. Let $r_i \geq 0$ be the number of observed complete event pairs for subject i. Then, $\delta_{ijk} = 1$ and $T_{ijk} = T^*_{ijk}$ for $k = 1, ..., r_i$ while $\delta_{i0(r_i+1)} = 1$ and $\delta_{i1(r_i+1)} = 0$ for $k = r_i + 1$. Accordingly, $T_{i0(r_i+1)} = T^*_{i0(r_i+1)}$ while $T_{i1(r_i+1)} = C_i$. In order to define observed gap times, first censoring C^w_{ijk} is defined where $C^w_{i0k} = C_i - T_{i1(k-1)}$ and $C^w_{i1k} = C_i - T_{i0k}$. Then, the observed gap times are $W_{ijk} = \min(W^*_{ijk}, C^w_{ijk})$.

Figure 2.2 illustrates the model. For instance, if alternating recurrent events are hospital admission and discharge, W_{i01} and W_{i11} are the first length of stay (LoS) and length of discharge (LoD), respectively. While the second LoS W_{i02} is uncensored, the last LoD W^*_{i12} is censored, so there is only one observed complete event pair. Note that since $W^*_{i12} > C^w_{i12}$, the observed gap time W_{i12} is C^w_{i12}.

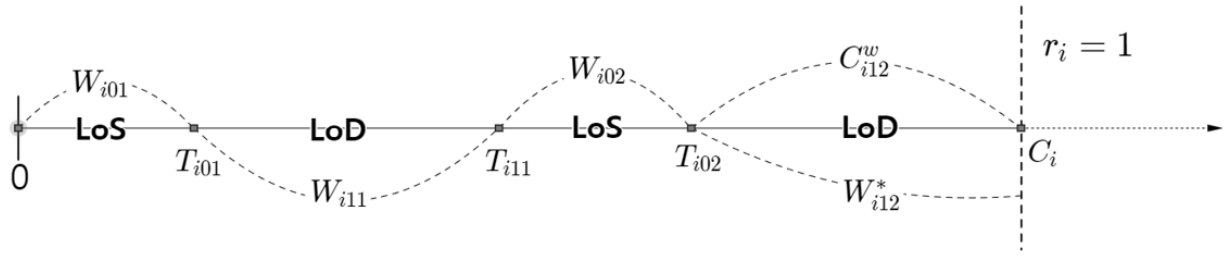

Figure 2.2: Alternating two-state process when $r_i = 1$

2.2. Joint transition intensity model

For each subject, there is one or more time-dependent or time-varying covariate for each gap time W_{ijk}, denoted $\boldsymbol{Z}_{ijk}$ where $\boldsymbol{Z}_{i0k} = \boldsymbol{Z}_{i0}(T^+_{i1(k-1)})$, $\boldsymbol{Z}_{i1k} = \boldsymbol{Z}_{i1}(T^+_{i0k})$ and T^+_{ijk} represents time immediately after T_{ijk}. That is, covariates for each gap time contain information on the subject promptly after the event occurring at the gap time origin. Note that the dimensions of the covariate may differ by event type. Also, given $\boldsymbol{Z}_{ijk}$, we assume that C_i is independent

of all event times.

Let $\boldsymbol{Z}_i = \{\boldsymbol{Z}_{ijk}, j = 0, 1, k = 1, ..., r_i + 1\}$ be the covariate history of subject i. Given $\boldsymbol{Z}_i$, we define the transition intensity model from j to $1 - j$ in terms of the alternating gap time sequence using a Cox proportional hazards model:

$$\lambda_{ijk}(w|\boldsymbol{Z}_i) = \lambda_{0j}(w) \exp(\boldsymbol{\beta}_j^T \boldsymbol{Z}_{ijk}),$$

where $\lambda_{0j}(\cdot)$ is the baseline transition intensity and $\boldsymbol{\beta}_j$ is the vector of regression coefficients. This model has two notable points: First, by defining the intensity model as a function of gap time instead of total event time, it accommodates recurrent events with non-negligible duration. Second, though the transition intensity function is defined for each occurrence, event type, and subject to consider covariates at each gap time, the baseline transition intensity function and regression coefficient vector are defined only for each event type. This approach is reasonable because event duration is generated for each event type, and covariates are usually common across subject and occurrence but vary by event type.

A potential problem with this model is that it does not permit correlation of outcomes within subjects, potentially resulting from unmeasured confounders. To address this issue, we adopt a model that includes individual frailty terms in the intensity linear predictor:

$$\lambda_{ijk}(w|\boldsymbol{Z}_i, \boldsymbol{u}_i) = \lambda_{0j}(w) \exp(\boldsymbol{\beta}_j^T \boldsymbol{Z}_{ijk} + u_{ij}) \tag{2.1}$$

Here, u_{i0} and u_{i1} are the individual transition frailties to the event type 1 and type 0, respectively. We call this the *joint transition intensity model (JTIM)*. To complete its specification, we assume that the vector of transition frailties $\boldsymbol{u}_i = (u_{i0}, u_{i1})^T$ follows a bivariate normal distribution $N(\boldsymbol{0}_2, \boldsymbol{\Sigma}_{2\times 2})$ where $\boldsymbol{\Sigma}_{2\times 2} = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{pmatrix}$ is a positive-definite variance-covariance matrix. Note that through the covariance term in $\boldsymbol{\Sigma}$, the dependency between event durations can be identified and this allows a more flexible model. Secondly, we assume that $\boldsymbol{u}_i$'s are mutually independent and $\boldsymbol{u}_i$ is independent of $\boldsymbol{Z}_i, \forall i$. Lastly, for $(j, k) \neq (p, q)$, W_{ijk}^* and W_{ipq}^* are independent given $\{\boldsymbol{u}_i, \boldsymbol{Z}_{ijk}, \boldsymbol{Z}_{ipq}\}$.

We combine the parameters in (2.1) in a vector denoted by $\boldsymbol{\xi} = (\lambda_{00}(\cdot), \lambda_{01}(\cdot), \boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T)^T$. Then, for subject i, the likelihood conditional on $\boldsymbol{u}_i$ and $\boldsymbol{Z}_i$ is

$$\begin{aligned} L_i(\boldsymbol{\xi}|\boldsymbol{u}_i, \boldsymbol{Z}_i) &= \prod_{j=0}^{1} \prod_{k=1}^{r_i+1} \{\lambda_{ijk}(W_{ijk}|\boldsymbol{Z}_i, \boldsymbol{u}_i)\}^{\delta_{ijk}} S_{ijk}(W_{ijk}|\boldsymbol{Z}_i, \boldsymbol{u}_i) \\ &= \prod_{j=0}^{1} \prod_{k=1}^{r_i+1} \{\lambda_{0j}(W_{ijk}) \exp(\boldsymbol{\beta}_j^T \boldsymbol{Z}_{ijk} + u_{ij})\}^{\delta_{ijk}} \{\exp(-\Lambda_{0j}(W_{ijk}) \exp(\boldsymbol{\beta}_j^T \boldsymbol{Z}_{ijk} + u_{ij}))\} \end{aligned}$$

where $S_{ijk}(\cdot|\boldsymbol{Z}_i, \boldsymbol{u}_i)$ is the conditional survival function which can be readily derived from (2.1) and $\Lambda_{0j}(w) = \int_0^w \lambda_{0j}(u)du$ is the cumulative baseline transition intensity [7]. Then, the joint likelihood for all subjects is $L = \prod_{i=1}^n L_i(\boldsymbol{\xi}|\boldsymbol{u}_i, \boldsymbol{Z}_i)$ and the marginal likelihood is the expectation of L with respect to $\boldsymbol{u}_i$, which is

$$\begin{aligned} L_m &= L(\boldsymbol{\xi}|\boldsymbol{Z}_i) \\ &= \int_{\boldsymbol{u}_i} \prod_{i=1}^{n} L_i(\boldsymbol{\xi}|\boldsymbol{u}_i, \boldsymbol{Z}_i) g(\boldsymbol{u}_i|\boldsymbol{\Sigma}) d\boldsymbol{u}_i \end{aligned} \tag{2.2}$$

where $g(\cdot|\boldsymbol{\Sigma})$ is the pdf of the bivariate normal distribution with mean $\mathbf{0}_2$ and variance $\boldsymbol{\Sigma}$.

2.3. Types of time-dependent covariate

Time-dependent covariates are classified into two broad classes [16]: External and internal. An *external* covariate is one whose values are determined independently of an event process, e.g. air pollution levels or age. Its future path is not affected by the occurrence of events although it can influence the rate of events over time. An *internal* covariate is one whose path is affected by survival status; in many cases, it requires the survival of the individual for its existence. Examples of internal covariates include biomarkers and other patient parameters that one can measure only if an individual is alive.

A Cox model can include both types of covariates. The distinction comes into play when constructing survival distributions. For example, if we have an internal covariate history

$\boldsymbol{Z}^H(t) = \{Z(u), 0 \leq u \leq t\}$, the conditional survival distribution

$$S(t|\boldsymbol{Z}^H(t)) = \Pr(T \geq t|\boldsymbol{Z}^H(t))$$

may not make sense because $\boldsymbol{Z}^H(t)$ was measured when an individual was alive at time t. For this reason, predicting survival with internal covariates is not straightforward though some internal covariates may explain survival well. One approach to accommodating internal covariates is to adopt a point process formulation and simulate simultaneously $N(t)$, the number of events until t, and $Z(t)$ ahead in time until $N(t) = 1$ [16].

In recurrent event modeling, the distinction is irrelevant because neither event is terminal. Therefore at a designated future time, regardless of whether an event occurs, a subject is still alive so that one can measure the covariates. Thus, $\boldsymbol{Z}_i$ can include both internal and external covariates. However, because alternating events create a history of past events, some internal covariates may be functions of the past event process itself. Examples of such internal covariates in recurrent event modeling include the cumulative number of type j events, the most recent gap time for type j events, or the number of type j events in the past (e.g. the number of 1-year previous admissions) [7]. Even if there are some missing values in covariates, the first example can be collected or generated if a subject merely knows how many events he has experienced. This example is the most common internal covariate used in recurrent event modeling. For readmission of patients with diabetes, for example, several studies describe a positive association between increasing length of a recent hospital stay and readmission risk [5][33][36].

2.4. Parameter estimation

We estimate the parameter of the JTIM in two stages: First, we adopt iterative estimating methods from Wang et al. [38] for the regression parameters $\boldsymbol{\beta}_j$, frailties $\boldsymbol{u}_i$, and variance-covariance matrix $\boldsymbol{\Sigma}$. Second, we estimate the baseline transition intensity $\lambda_{0j}(\cdot)$

nonparametrically *via* the Breslow estimator and B-spline regression.

Under the general JTIM (2.1), there is no closed form for the marginal likelihood (2.2). To overcome this, Wang et al. derived an approximate marginal likelihood. Let $\boldsymbol{u} = (\boldsymbol{u}_1^T, \boldsymbol{u}_2^T, ..., \boldsymbol{u}_n^T)^T$ be a vector stacking frailties of all subjects, which then follows a mean-zero multivariate normal distribution MVN$(\boldsymbol{0}_{2n}, \boldsymbol{D}_{2n\times 2n})$. Here $\boldsymbol{D} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_{n\times n}$ is a block-diagonal matrix, where $\otimes$ is a Kronecker product and $\boldsymbol{I}_{n\times n}$ is a n by n identity matrix. Also, let $\boldsymbol{R}_{ij} = (0_{(1)}, ..., 1_{(2i-1+j)}, ..., 0_{(2n)})$ be frailty design vectors to indicate presence of u_{ij} in the $(2i-1+j)$th entry of $\boldsymbol{u}$. Then, (2.1) becomes

$$\lambda_{ijk}(w|\boldsymbol{Z}_i, \boldsymbol{R}_{ij}, \boldsymbol{u}_i) = \lambda_{0j}(w)\exp(\boldsymbol{\beta}_j^T \boldsymbol{Z}_{ijk} + \boldsymbol{u}^T \boldsymbol{R}_{ij}),$$

and the joint likelihood function from the n subjects and their frailties is

$$L(\boldsymbol{\xi}, \boldsymbol{u}) = |\boldsymbol{D}|^{-1/2}\exp(-\boldsymbol{K}(\boldsymbol{u})), \tag{2.3}$$

where

$$\begin{aligned}\boldsymbol{K}(\boldsymbol{u}) = \ & \frac{1}{2}\boldsymbol{u}^T\boldsymbol{D}^{-1}\boldsymbol{u} + \sum_{i=1}^{n}\sum_{j=0}^{1}\sum_{k=1}^{r_i+1}\Lambda_{0j}(W_{ijk})\exp(\boldsymbol{\beta}_j^T\boldsymbol{Z}_{ijk} + \boldsymbol{u}^T\boldsymbol{R}_{ij}) \\ & - \delta_{ijk}\{\boldsymbol{\beta}_j^T\boldsymbol{Z}_{ijk} + \boldsymbol{u}^T\boldsymbol{R}_{ij} + \log(\lambda_{0j}(W_{ijk}))\}.\end{aligned}$$

The authors approximated $\boldsymbol{K}(\boldsymbol{u})$ through a second-order Taylor expansion denoted by $\hat{\boldsymbol{K}}(\boldsymbol{u})$, plugged it into (2.3), and integrated it to obtain an approximate marginal log-likelihood:

$$l_m = \log(L_m) \approx -\frac{1}{2}\log|\boldsymbol{D}| - \boldsymbol{K}(\tilde{\boldsymbol{u}}) - \frac{1}{2}\log|\boldsymbol{K}_2(\tilde{\boldsymbol{u}})|$$

where $\tilde{\boldsymbol{u}}$ is the solution of $\frac{\partial \boldsymbol{K}(\boldsymbol{u})}{\partial \boldsymbol{u}} = 0$ and $\boldsymbol{K}_2(\boldsymbol{u})$ is second derivative of $\boldsymbol{K}(\boldsymbol{u})$.

Another challenge is simultaneous estimation of the baseline transition intensity function. Wang et al. found that considering the likelihood including baseline intensity function critically slows down the computations, and direct replacement of $\Lambda_{0j}(\cdot)$ with its correspond-

ing Breslow estimator overestimates the diagonal entries in $\boldsymbol{\Sigma}$. To avoid this problem, they utilized the penalized partial log-likelihood. Note that the function $\boldsymbol{K}(\boldsymbol{u})$ is the negative penalized joint log-likelihood, which can be factored as

$$\boldsymbol{K}(\boldsymbol{u}) = -\text{PPLL} - h\left(\lambda_{00}(\cdot), \lambda_{01}(\cdot), \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{u}\right),$$

where PPLL indicates the penalized partial log-likelihood and $h(\cdot)$ contains all other terms accompanying $\lambda_{0j}(\cdot)$. Ripatti and Palmgren [29] demonstrated that the information loss from ignoring $h(\cdot)$ is negligible, so that one can estimate $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ and $\boldsymbol{u}$ solely based on the PPLL terms. They also presented that neglecting $h(\cdot)$ and estimation of the baseline transition intensity functions substantially accelerates the computation.

We estimate the regression coefficients and variance components by iterating two steps, an inner loop and an outer loop. First, set initial values for $\boldsymbol{\beta}_j, \boldsymbol{u}, \boldsymbol{\Sigma}$. In the inner loop, $\boldsymbol{\beta}_j$ and $\boldsymbol{u}$ are estimated through the PPLL using a Newton-Raphson algorithm with initial values. Note that $\boldsymbol{\Sigma}$ is considered as a known parameter which is an initial value in the beginning and later, an estimate derived from the previous outer loop. Using a recursive estimating method, $\boldsymbol{\Sigma}$ is estimated in the outer loop through an approximate marginal profile log-likelihood given the estimates of $\boldsymbol{\beta}_j$ and $\boldsymbol{u}$ from the previous inner loop.

Before we examine the estimation of baseline transition intensity functions $\lambda_{0j}(\cdot)$, we emphasize that our interest is in the length of the unknown gap time of the last event, which depends on survivor functions derived from (2.1). Thus it is enough to estimate the cumulative baseline transition intensity $\Lambda_{0j}(\cdot)$ rather than $\lambda_{0j}(\cdot)$.

We apply the Breslow estimator [4][21] for $\Lambda_{0j}(\cdot)$ as follows:

$$\Lambda_{0j}^{B}(w) = \sum_{i=1}^{n} \sum_{k=1}^{r_i+1} \frac{I(W_{ijk} \leq w)\delta_{ijk}}{\sum_{p=1}^{n} \sum_{q=1}^{r_p+1} I(W_{pjq} \geq W_{ijk}) \exp(\hat{\boldsymbol{\beta}}_j^T \boldsymbol{Z}_{pjq} + \hat{u}_{pj})},$$

which indicates that in terms of a gap time w, the cumulative baseline transition intensity is estimated conditional on $\hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{u}}$, which are in turn estimated in the previous estimation part. Once we obtain $\hat{\Lambda}_{0j}^{B}(w)$ values corresponding to the observed W_{ijk} values, B-spline

regression [13] is implemented to estimate the function $\Lambda_{0j}(\cdot)$. We use a linear spline to maintain the non-decreasing property of $\Lambda_{0j}(\cdot)$. In the left boundary interval, nevertheless, the linear fitting sometimes produces improper estimates, as Figure 2.3 demonstrates. On the left plot, the linear spline causes negative estimates for observations close to 0 while the linear spline on the right plot over-estimates the observations close to 0. To prevent this, a cubic or higher order spline is applied in the left boundary interval, resolving the issue in both plots.

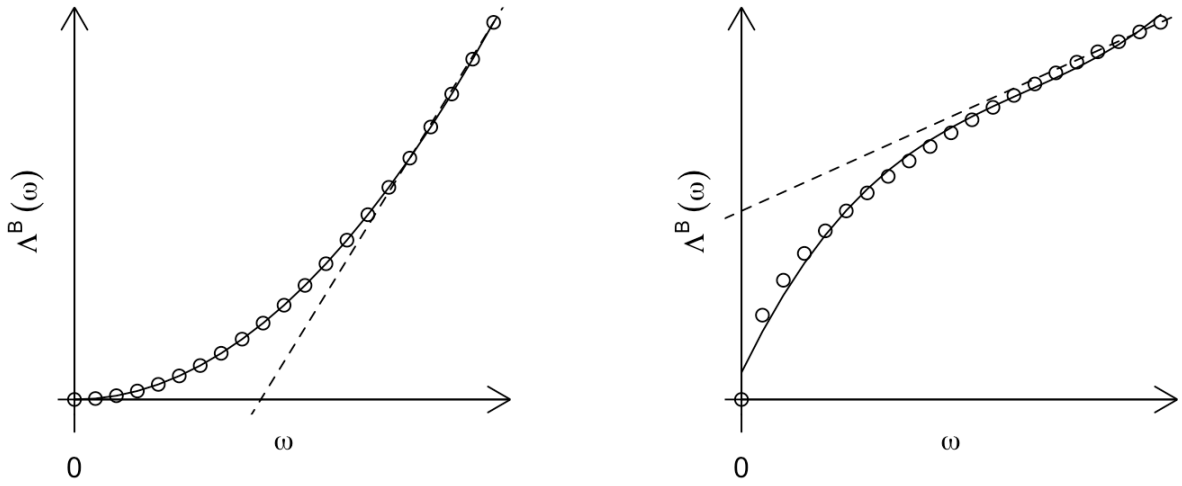

Figure 2.3: Spline examples in left boundary interval. The solid and dashed line shows cubic and linear spline, respectively.

In order to select proper knots, moreover, we use the A-splines (adaptive splines) approach including automatic knot selection by Goepp et al. [11]. The approach sets a large number of initial knots and fits B-spline regression with an iterative penalized likelihood called the adaptive ridge to sequentially remove the unnecessary knots. We select the model achieving the best bias-variance tradeoff by the Bayesian information criterion.

CHAPTER 3

Dynamic Prediction of Transition Risk

3.1. Predictive transition probability considering process history

Assume that we have estimated a JTIM with existing subjects, and a new subject i has arrived. We have information about his covariate history and recurrent events up to a censoring time. We seek to use this information to estimate the probability of transitioning to event type 0 within τ after censoring time C_i. That is, we estimate the probability that unobserved true gap time $W^*_{i1(r_i+1)}$ is less than or equal to observed censored gap time $C^w_{i1(r_i+1)} + \tau$ where τ is the *window of prediction*. We express the transition probability as

$$\Pr(W^*_{i1(r_i+1)} \leq C^w_{i1(r_i+1)} + \tau | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1}), \tag{3.1}$$

where $H_i = \{W_{i11}, W_{i12}, ..., W_{i1r_i}\}$ is the subject's gap-time history of event type 1. Note that because the last gap time is always censored, $W_{i1(r_i+1)} = C^w_{i1(r_i+1)}$. To obtain (3.1), we fix estimates of $\boldsymbol{\xi}, \boldsymbol{\Sigma}$ from the fitted model, but the subject-level transition frailty u_{i1} is unknown for this new subject. To address this, we propose estimating the marginal probability of (3.1), which we obtain by integrating out the frailty term

$$\begin{aligned} &\Pr(W^*_{i1(r_i+1)} \leq C^w_{i1(r_i+1)} + \tau | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) \\ &= \int_{u_{i1}} \Pr(W^*_{i1(r_i+1)} \leq C^w_{i1(r_i+1)} + \tau | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1}) \\ &\quad \times f_{u_{i1}}(u_{i1} | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) du_{i1}, \end{aligned} \tag{3.2}$$

where the second integrand is the posterior density of u_{i1}.

The first integrand is derived as follows;

$$\begin{aligned}
&\Pr(W^*_{i1(r_i+1)} \le C^w_{i1(r_i+1)} + \tau | W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1}) \\
&= \frac{\Pr(C^w_{i1(r_i+1)} \le W^*_{i1(r_i+1)} \le C^w_{i1(r_i+1)} + \tau, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})}{\Pr(W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})} \\
&= \frac{\Pr(C^w_{i1(r_i+1)} \le W^*_{i1(r_i+1)} \le C^w_{i1(r_i+1)} + \tau, H_i | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})}{\Pr(W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})}.
\end{aligned}$$

Recall that gap times are assumed to be mutually independent given $u_{i1}, \boldsymbol{Z}_i$. Then,

$$\begin{aligned}
&\frac{\Pr(C^w_{i1(r_i+1)} \le W^*_{i1(r_i+1)} \le C^w_{i1(r_i+1)} + \tau, H_i | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})}{\Pr(W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})} \\
&= \frac{\Pr(C^w_{i1(r_i+1)} \le W^*_{i1(r_i+1)} \le C^w_{i1(r_i+1)} + \tau | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1}) \Pr(H_i | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})}{\Pr(W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)} | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1}) \Pr(H_i | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})} \\
&= \frac{\Pr(C^w_{i1(r_i+1)} \le W^*_{i1(r_i+1)} \le C^w_{i1(r_i+1)} + \tau | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})}{\Pr(W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)} | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})} \\
&= \frac{S_{i1(r_i+1)}(C^w_{i1(r_i+1)} | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1}) - S_{i1(r_i+1)}(C^w_{i1(r_i+1)} + \tau | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})}{S_{i1(r_i+1)}(C^w_{i1(r_i+1)} | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1})},
\end{aligned} \tag{3.3}$$

where $S_{i1(r_i+1)}(\cdot)$ is the survivor function for occurrence (r_i+1) of type 1 for subject i.

For the posterior density of u_{i1}, because u_{i1} and $\boldsymbol{Z}_i$ are assumed to be independent, and $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}$ are fixed,

$$\begin{aligned}
&f_{u_{i1}}(u_{i1} | W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) \\
&= \frac{f_{u_{i1} C^w_{i1(r_i+1)} H_i}(u_{i1}, W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma})}{f_{C^w_{i1(r_i+1)} H_i}(W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma})} \\
&= \frac{f_{C^w_{i1(r_i+1)} H_i}(W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i | u_{i1}, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) f_{u_{i1}}(u_{i1} | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma})}{\displaystyle\int_{u_{i1}} f_{C^w_{i1(r_i+1)} H_i}(W^*_{i1(r_i+1)} \ge C^w_{i1(r_i+1)}, H_i | u_{i1}, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) f_{u_{i1}}(u_{i1} | \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) du_{i1}}.
\end{aligned}$$

Here, $f_{u_{i1}}(\cdot|\sigma_1)$ is the normal density with mean 0 and standard deviation σ_1. Next we derive

$$\begin{aligned}
&\frac{f_{C^w_{i1(r_i+1)}H_i}(W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i|u_{i1}, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) f_{u_{i1}}(u_{i1}|\boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma})}{\int_{u_{i1}} f_{C^w_{i1(r_i+1)}H_i}(W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i|u_{i1}, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) f_{u_{i1}}(u_{i1}|\boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) du_{i1}} \\
&= \frac{f_{C^w_{i1(r_i+1)}}(W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}|u_{i1}, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) f_{H_i}(H_i|u_{i1}, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}} f_{C^w_{i1(r_i+1)}}(W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}|u_{i1}, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) f_{H_i}(H_i|u_{i1}, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}) f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}} \\
&= \frac{\prod_{k=1}^{r_i+1} \left(-\left.\frac{\partial S_{i1k}(w)}{\partial w}\right|_{w=W_{i1k}} \right)^{\delta_{i1k}} \left(S_{i1k}(W_{i1k}) \right)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}} \prod_{k=1}^{r_i+1} \left(-\left.\frac{\partial S_{i1k}(w)}{\partial w}\right|_{w=W_{i1k}} \right)^{\delta_{i1k}} \left(S_{i1k}(W_{i1k}) \right)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}} \\
&= \frac{\prod_{k=1}^{r_i+1} \left(-S'_{i1k}(W_{i1k}) \right)^{\delta_{i1k}} \left(S_{i1k}(W_{i1k}) \right)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}} \prod_{k=1}^{r_i+1} \left(-S'_{i1k}(W_{i1k}) \right)^{\delta_{i1k}} \left(S_{i1k}(W_{i1k}) \right)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}},
\end{aligned} \tag{3.4}$$

where $\left.\frac{\partial S_{i1k}(w)}{\partial w}\right|_{w=W_{i1k}} = S'_{i1k}(W_{i1k})$ for simplicity. Combining (3.3) and (3.4) gives

$$\begin{aligned}
&\Pr(W^*_{i1(r_i+1)} \leq C^w_{i1(r_i+1)} + \tau | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}, u_{i1}) \\
&= \int_{u_{i1}} \frac{\frac{S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) - S_{i1(r_i+1)}(C^w_{i1(r_i+1)} + \tau)}{S_{i1(r_i+1)}(C^w_{i1(r_i+1)})} \prod_{k=1}^{r_i+1} \left(-S'_{i1k}(W_{i1k}) \right)^{\delta_{i1k}} \left(S_{i1k}(W_{i1k}) \right)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}} \prod_{k=1}^{r_i+1} \left(-S'_{i1k}(W_{i1k}) \right)^{\delta_{i1k}} \left(S_{i1k}(W_{i1k}) \right)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}} du_{i1}.
\end{aligned}$$

When $k = r_i + 1$, the observation of event type 1 is censored, i.e. $\delta_{i1(r_i+1)} = 0$,

so $\Big(S_{i1k}(W_{i1k})\Big)^{1-\delta_{i1k}} = S_{i1(r_i+1)}\Big(C^w_{i1(r_i+1)}\Big)$, giving

$$\int_{u_{i1}} \frac{\frac{S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) - S_{i1(r_i+1)}(C^w_{i1(r_i+1)}+\tau)}{S_{i1(r_i+1)}(C^w_{i1(r_i+1)})} \prod_{k=1}^{r_i+1}\Big(-S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}}\Big(S_{i1k}(W_{i1k})\Big)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}} \prod_{k=1}^{r_i+1}\Big(-S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}}\Big(S_{i1k}(W_{i1k})\Big)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}} du_{i1}$$

$$= \int_{u_{i1}} \frac{\frac{S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) - S_{i1(r_i+1)}(C^w_{i1(r_i+1)}+\tau)}{S_{i1(r_i+1)}(C^w_{i1(r_i+1)})} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) \prod_{k=1}^{r_i}\Big(-S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) \prod_{k=1}^{r_i}\Big(-S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}} du_{i1}$$

$$= \frac{\int_{u_{i1}} \Big[S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) - S_{i1(r_i+1)}(C^w_{i1(r_i+1)}+\tau)\Big] \prod_{k=1}^{r_i}\Big(-S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}}{\int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) \prod_{k=1}^{r_i}\Big(-S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}}. \tag{3.5}$$

Here,

$$\begin{aligned} S_{i1k}(W_{i1k}) &= \exp\big(-\Lambda_{i1k}(W_{i1k})\big) \\ &= \exp\big(-\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\big) \end{aligned}$$

and

$$\begin{aligned} -S'_{i1k}(W_{i1k}) &= \lambda_{i1k}(W_{i1k})S_{i1k}(W_{i1k}) \\ &= \lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\exp\big(-\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\big). \end{aligned}$$

We can interpret (3.5) as the ratio of marginal likelihoods; the denominator shows the marginal likelihood of the gap time history, and the numerator represents the marginal likelihood of gap time history including the next event occurring between $C^w_{i1(r_i+1)}$ and $C^w_{i1(r_i+1)}+\tau$. If follow-up is censored immediately after the gap time origin of $W^*_{i1(r_i+1)}$, then $C^w_{i1(r_i+1)}$ would be the smallest time unit bigger than 0.

Note that our predictive transition probability directly accommodates the cumulative number of past type 1 events as an internal time-varying covariate. Expanding the posterior distribution (3.4),

$$
\begin{aligned}
&\frac{\prod_{k=1}^{r_i+1}\Big(-S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}}\Big(S_{i1k}(W_{i1k})\Big)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}}\prod_{k=1}^{r_i+1}\Big(-S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}}\Big(S_{i1k}(W_{i1k})\Big)^{1-\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}} \\
&= \frac{\prod_{k=1}^{r_i+1}\Big[(\lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1}))^{\delta_{i1k}}\exp\big(-\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\big)\Big] f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}}\prod_{k=1}^{r_i+1}\Big[(\lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1}))^{\delta_{i1k}}\exp\big(-\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\big)\Big] f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}} \\
&= \frac{\prod_{k=1}^{r_i+1}\Big[e^{u_{i1}\delta_{i1k}}\exp\big(-\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\big)\Big] f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}}\prod_{k=1}^{r_i+1}\Big[e^{u_{i1}\delta_{i1k}}\exp\big(-\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\big)\Big] f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}} \\
&= \frac{\exp\Big(\sum_{k=1}^{r_i+1}u_{i1}\delta_{i1k}-\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\Big) f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}}\exp\Big(\sum_{k=1}^{r_i+1}u_{i1}\delta_{i1k}-\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\Big) f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}} \\
&= \frac{\exp\Big(r_iu_{i1}-\sum_{k=1}^{r_i+1}\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\Big) f_{u_{i1}}(u_{i1}|\sigma_1)}{\int_{u_{i1}}\exp\Big(r_iu_{i1}-\sum_{k=1}^{r_i+1}\Lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}+u_{i1})\Big) f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}}.
\end{aligned}
$$

We can notice that through r_iu_{i1} term, the individual frailty transiting to event type 0 is considered in proportion to the number of past events.

3.2. Approximate transition probability

A drawback of (3.5) is that it has no closed form, so that numerical integrations are required for both the numerator and denominator. To avoid numerical integration and

simplify computing, we propose an approximation of the marginal transition probability by Taylor expansion. The brief procedure is as follows (for details, see Appendix A.1):

By Taylor expansion, $e^{u_{i1}} = 1 + u_{i1} + \frac{u_{i1}^2}{2!} + o(u_{i1}^2)$ as $u_{i1} \to 0$. Letting $\eta(W_{ijk}) = \Lambda_{01}(W_{ijk}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k})$, then

$$\begin{aligned} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) &= \exp\Big(- \Lambda_{01}(C^w_{i1(r_i+1)}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k} + u_{i1})\Big) \\ &= \exp\Big(- \eta(C^w_{i1(r_i+1)}) e^{u_{i1}}\Big) \\ &\approx \exp\Big(- \eta(C^w_{i1(r_i+1)})(1 + u_{i1} + u_{i1}^2/2)\Big) \end{aligned}$$

and

$$\begin{aligned} \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} &= \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k} + u_{i1}) \exp\Big(- \Lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k} + u_{i1})\Big) \\ &= \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}) e^{u_{i1}} \exp\Big(- \eta(W_{i1k}) e^{u_{i1}}\Big) \\ &= \exp\Big(r_i u_{i1} - e^{u_{i1}} \sum_{k=1}^{r_i} \eta(W_{i1k})\Big) \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}) \\ &\approx \exp\Big(r_i u_{i1} - (1 + u_{i1} + u_{i1}^2/2) \textstyle\sum_{k=1}^{r_i} \eta(W_{i1k})\Big) \displaystyle\prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}). \end{aligned}$$

We plug these into (3.5) and obtain closed forms for the numerator and denominator. The numerator is

$$\int_{u_{i1}} \Big[S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) - S_{i1(r_i+1)}(C^w_{i1(r_i+1)}+\tau)\Big] \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$= \int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$- \int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}+\tau) \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$\approx \exp\left(\frac{\sigma_1^2 \Big(r_i - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)})\Big)^2}{2\Big(1+\sigma_1^2 \sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)})\Big)} - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)}) \right)$$

$$\times \Big(1+\sigma_1^2 \textstyle\sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)})\Big)^{-\frac{1}{2}} \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k})$$

$$- \exp\left(\frac{\sigma_1^2 \Big(r_i - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)}+\tau)\Big)^2}{2\Big(1+\sigma_1^2 \sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)}+\tau)\Big)} - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)}+\tau) \right)$$

$$\times \Big(1+\sigma_1^2 \textstyle\sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)}+\tau)\Big)^{-\frac{1}{2}} \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}),$$

and the denominator is

$$\int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$\approx \exp\left(\frac{\sigma_1^2 \Big(r_i - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)})\Big)^2}{2\Big(1+\sigma_1^2 \sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)})\Big)} - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)}) \right)$$

$$\times \Big(1+\sigma_1^2 \textstyle\sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)})\Big)^{-\frac{1}{2}} \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}).$$

CHAPTER 4

Simulation Studies

4.1. Simulated data and models compared

We generated $N = 500$ sets of patient admission and discharge records during 3 years where the time unit is day. We set every patient to be censored 3 years after entering the first admission; thus either the last admission or discharge gap time can be censored.

We include two external covariates, gender and age. Patients were randomly assigned to male or female, with age between 1 and 65 at entry. Regression coefficients for the covariates are $\boldsymbol{\beta}_0 = (-0.05, -0.01)^T$ for admission and $\boldsymbol{\beta}_1 = (-0.5, 0.001)^T$ for discharge event. Taking male as the reference category, females are less likely to be discharged earlier and readmitted. Also, older patients are less likely to be discharged early and more likely to be readmitted than younger patients.

Transition frailties to be discharged and readmitted follow a normal distribution with mean 0 and variances σ_0^2, σ_1^2, respectively. We assume that the frailties are negatively correlated with $\rho = -0.5$, implying that LoS is inversely associated with LoD. Lastly, we take both baseline transition intensities to be constant.

Table 4.1 presents some example data. There are records of two patients, with the first one a young male who has 3 past admission and discharge records with the last LoD censored. The second patient is a middle-aged female who has no admission records after discharge from her first admission.

ID	LoS	LoD	Gender	Age	Censoring
1	2	254	M	28	N
1	2	684	M	28	N
1	19	134	M	30	Y
2	5	1090	F	52	Y

Table 4.1: Simulated data example

We compared the prediction performance of 4 different models were adapted: JTIM; a logistic mixed model (LMM); generalized estimating equations (GEE); and logistic regression (LR). To create predictions on future patients in LMM, model parameters are readily estimated, but still we need to estimate patient-specific random effects. We propose deriving the posterior density of the random effects using all available past records to date and estimating the posterior mean. Specifically, given the unnormalized patient-specific posterior density, we estimate the posterior mean of the random effect, and create a predicted value using the estimated LMM parameters and the posterior mean. Details are as follows:

For hospitalization k of patient i, let y_{ik} be the value of a binary target variable indicating whether a readmission occurs within τ days. Then, the distribution of y_{ik} is

$$Y_{ik} \sim Bin(1, \pi_{ik}),$$

where $\pi_{ik} = \Pr(Y_{ik} = 1)$ is the probability that a readmission occurs within τ days after hospitalization k. Also, $\boldsymbol{Z}_{ik}$ denotes a predictor vector including gender and age. For subject i, let u_i be a normally distributed patient-specific random effect with zero mean and constant variance σ^2. We assume that the following logistic mixed model:

$$E(Y_{ik}|\boldsymbol{\beta}, \boldsymbol{Z}_{ik}, u_i) = \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{Z}_{ik} + u_i)}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_{ik} + u_i)}, \quad i = 1, ..., N, \ k = 1, ..., r_i + 1,$$

where $\boldsymbol{\beta}$ is a coefficient vector for gender and age. Here, we assume that random effects are independent between patients. One can readily obtain the maximum likelihood estimate

(MLE) of $\boldsymbol{\xi} = \{\boldsymbol{\beta}, \sigma^2\}$ by applying optimization techniques to a marginal likelihood calculated by Gaussian quadrature. Once we have $\hat{\boldsymbol{\xi}}$, the posterior distribution of u_i at the k-th hospitalization is

$$f_{u_i}(u_i|\boldsymbol{y}_i^{(-k)}, \boldsymbol{\xi}) = \frac{\Pr(\boldsymbol{y}_i^{(-k)}|u_i, \boldsymbol{\xi}) f_{u_i}(u_i|\boldsymbol{\xi})}{\int_{u_i} \Pr(\boldsymbol{y}_i^{(-k)}|u_i, \boldsymbol{\xi}) f_{u_i}(u_i|\boldsymbol{\xi}) du_i} \tag{4.1}$$

where $\boldsymbol{y}_i^{(-k)}$=$\{y_{i1}, ..., y_{i,k-1}\}$, $\Pr(\boldsymbol{y}_i^{(-k)}|u_i, \boldsymbol{\xi}) = \prod_{l=1}^{k-1} \Pr(y_{il}|u_i, \boldsymbol{\xi})$, and $f(u_i|\boldsymbol{\xi})$ is the normal pdf with mean 0 and standard deviation σ. Note that $k = 1$ implies that there is no history, so (4.1) is equal to the prior distribution $f_{u_i}(u_i|\boldsymbol{\xi})$. The posterior mean is

$$E(u_i|\boldsymbol{y}_i^{(-k)}, \boldsymbol{\xi}) = u_{ik} = \int_{-\infty}^{\infty} u_i f_{u_i}(u_i|\boldsymbol{y}_i^{(-k)}, \boldsymbol{\xi}) du_i.$$

We obtain the posterior mean by numerical integration. With $\hat{\boldsymbol{\xi}}$ and $\hat{u}_{ik}$, the estimated predictive probability that a readmission occurs within τ days after the k-th hospitalization of patient i is

$$\widehat{\Pr}(Y_{ik} = 1|\hat{\boldsymbol{\beta}}, \boldsymbol{Z}_{ik}, \hat{u}_{ik}) = \frac{\exp(\hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_{ik} + \hat{u}_{ik})}{1 + \exp(\hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_{ik} + \hat{u}_{ik})}.$$

As mentioned, JTIM automatically captures the cumulative number of past admissions. In contrast, GEE and LM use information from the latest admission and discharge only. Even in LMM, though the patient random effect is obtained by using posterior distribution, LMM does not use the number of past admissions directly when predicting the transition probability. To compare all models as fairly as possible, we add

$$\text{average annual \# of past admissions} = \frac{\text{cumulative \# of past admissions}}{\text{follow-up period (year)}}$$

as an *ad hoc* covariate for models other than JTIM. With 3 covariates including gender, age, and the average annual number of past admissions, we used GEE with the AR(1) correlation structure. Also, because LR does not account for the correlation between records within a patient, we randomly select one record from each patient's hospitalizations to fit the model. Lastly, patients in simulated data are always randomly divided into an 80% training set and

20% test set.

4.2. Prediction performance under different numbers of past records

Recall that r_i denotes the number of complete event pairs of patient i. Intuitively, the more past records a patient has, the better will JTIM predict compared to other methods, because the predictive probability will reflect an increasingly concentrated frailty distribution. Thus, we can expect that relative prediction performance will depend on the number of past records.

In this section, we have implemented simulations under different values of median r_i ($\text{med}(r_i)$). Using the simulated data above, first we set $\text{med}(r_i)$ equal to 0, 2, 4. For instance, if $\text{med}(r_i) = 0$, half or more of the patients have no complete admission/discharge pairs.

For each setting, we ran $M = 1,000$ Monte-Carlo simulations. We fitted the models with the training set and validated the prediction performance with the test set. We set variances of the frailties to be $\sigma_0^2 = \sigma_1^2 = 0.5$. The window of prediction τ was fixed at 30 days, which gives readmission rates from 18%−25% across $\text{med}(r_i)$. When predicted, we assumed that each patient is censored immediately after entering discharge status, i.e. the last censored gap time $C^w_{i1r_i+1}$ is the smallest time unit, one day.

Prediction performance was evaluated by three measures: Area under the receiver operating characteristic (ROC) curve (AUC); sensitivity — the probability that admissions with a subsequent 30-day readmission are correctly identified; and win rate — the fraction of 1,000 Monte-Carlo simulations where JTIM has higher performance than a comparator. We calculated the win rate separately for AUC and sensitivity.

We prefer to present sensitivity for the following reason: Imagine there are currently two patients hospitalized with diabetes. We know that one will be readmitted within 30 days after discharge and the other will not. If a prediction model predicted incorrectly for both patients, the one who will be readmitted would be affected badly because he would have

foregone any chance to avoid readmission. By contrast, the incorrect prediction would not cause significant damage to a subject who is not readmitted. Thus, we believe that high sensitivity in this context is generally more important than high specificity.

There are sensitivities as many as there are potential prediction thresholds. To select a reasonable threshold and calculate a sensitivity as a measure, we suggest following steps:

1. From the data generated in each simulation, obtain the observed 30-day readmission rate $\alpha \in [0, 1]$.

2. Between predictive transition probabilities for 30-day readmission for every discharge events, choose the $100(1 - \alpha)$th quantile as the threshold. For example, if the observed readmission rate is 0.1, then use the 90th centile between predictive transition probabilities as the prediction threshold.

If all admissions with a subsequent 30-day readmission have predictive probabilities higher than the threshold, the sensitivity is 1. However, it would be less than 1 if any of them has a predictive probability below the threshold.

Table 4.2 shows the AUC and sensitivity under varying $\text{med}(r_i)$. Overall, the JTIM shows better prediction performance than the other models in terms of both AUC and sensitivity. Furthermore, we can find some trends through the results. When $\text{med}(r_i) = 0$, JTIM and the other models perform similarly. However, greater $\text{med}(r_i)$ produces greater mean differences and win rates favoring JTIM.

When comparing the results of the other 3 models, LMM and GEE show similar results in AUC, but LMM produces better sensitivity when $\text{med}(r_i)$ increases. Though both models accommodate within-patient correlation, the results proves that considering information of past records through the posterior distribution contributes to higher sensitivity of LMM. LR, which ignores all but a single admission/discharge pair, performs less well than the other methods.

		$r_i = 0$		$r_i = 2$		$r_i = 4$	
		Mean (se)	Win rate	Mean (se)	Win rate	Mean (se)	Win rate
JTIM	AUC	.655 (.061)	·	.670 (.044)	·	.680 (.035)	·
	Sens.	.323 (.082)	·	.402 (.060)	·	.488 (.055)	·
LMM	AUC	.635 (.061)	.67	.644 (.044)	.86	.645 (.040)	.94
	Sens.	.306 (.081)	.74	.378 (.062)	.80	.460 (.053)	.83
GEE	AUC	.633 (.063)	.69	.640 (.046)	.88	.640 (.039)	.95
	Sens.	.299 (.082)	.75	.363 (.059)	.86	.445 (.051)	.91
LR	AUC	.627 (.062)	.70	.634 (.047)	.90	.630 (.041)	.98
	Sens.	.293 (.081)	.77	.357 (.060)	.90	.440 (.050)	.95

Table 4.2: AUC and sensitivity of JTIM and other models when median of r_i, the number of observed complete admission and discharge pairs, is varying.

4.3. Prediction performance under different variances of transition frailty to be readmitted

σ_1^2 is the variance of the frailty for transition to the admission state. As it increases, there will be more heterogeneity between patients, leading to increased variance of LoD. Hence, we can anticipate that models addressing the heterogeneity between patients can capture the difference in readmission intensity between patients and make more accurate predictions. In this section, we implemented simulations under different values of σ_1^2. Using the simulated data above, we varied $\sigma_1^2 = 0.25, 0.5, 0.75$ to see the effect of frailty variance on prediction performance.

For each setting, we ran $M = 1,000$ simulations. We fitted the models with the training set and validated prediction performance with the test set. The window of prediction τ was fixed at 30 days, leading to readmission rates from 18%−23%. For these parameters, med(r_i) is always around 2. When predicted, again we assumed that each patient is censored immediately after entering discharge status.

Table 4.3 shows AUC and sensitivity for varying σ_1^2. Patterns are similar to those in Table 4.2. First, JTIM always produces better performance than the other models in terms of both AUC and sensitivity. Also, as σ_1^2 increases, the mean advantage of JTIM over the other models generally increases.

The results suggest that our approach works better for higher heterogeneity between patients. Indeed, during simulations, the (10, 25, 50, 75, 90) centiles of r_i with $\sigma_1^2 = 0.25$ are (0,1,2,4,5) on average; with $\sigma_1^2 = 0.75$ it is (0,1,2,5,7). This is because with fixed $\text{med}(r_i) = 2$, some patients have more complete event pairs. In other words, increasing σ_1^2 produces higher transition frailties to admission, and patients with those higher transition frailties have shorter LoD and therefore more pairs. These patients with more records contribute to better performance of JTIM.

When comparing the results of the other 3 models, again LMM and GEE shows similar results in AUC, but LMM produces better sensitivity for increasing σ_1^2. LR produces similar or worse results across the board.

		$\sigma_1^2 = 0.25$		$\sigma_1^2 = 0.5$		$\sigma_1^2 = 0.75$	
		Mean (se)	Win rate	Mean (se)	Win rate	Mean (se)	Win rate
JTIM	AUC	.624 (.045)	·	.670 (.044)	·	.691 (.035)	·
	Sens.	.294 (.059)	·	.401 (.060)	·	.480 (.055)	·
LMM	AUC	.612 (.047)	.73	.644 (.044)	.86	.660 (.040)	.93
	Sens.	.274 (.060)	.74	.373 (.062)	.80	.450 (.053)	.84
GEE	AUC	.612 (.045)	.78	.640 (.046)	.88	.650 (.039)	.96
	Sens.	.270 (.059)	.75	.359 (.061)	.86	.431 (.051)	.91
LR	AUC	.598 (.043)	.80	.635 (.047)	.90	.645 (.041)	.96
	Sens.	.263 (.058)	.80	.353 (.060)	.90	.427 (.050)	.92

Table 4.3: AUC and sensitivity of JTIM and other models when σ_1^2 is varying.

4.4. Prediction performance under different readmission rates

This section examines how different readmission rates affect the prediction performance of JTIM compared to the other models. We set readmission rates at 10%, 30%, and 50%, corresponding to windows of prediction around 30, 90, and 180 days. Here, $\text{med}(r_i)$ and σ_1^2 are fixed at 2 and 0.5, respectively. As in the previous section, we ran $M = 1,000$ Monte-Carlo simulations for each setting.

Table 4.4 illustrates the results. Similar to Table 4.2, JTIM produces superior performance across all settings. Following JTIM, LMM shows better performance than GEE and LR. With a few exceptions, as the readmission rate increases, the win rate also increases in both AUC and sensitivity. LR improves substantially as the readmission rate increases.

		Readmission rate = 10%		Readmission rate = 30%		Readmission rate = 50%	
		Mean (se)	Win rate	Mean (se)	Win rate	Mean (se)	Win rate
JTIM	AUC	.660 (.050)	·	.680 (.042)	·	.691 (.038)	·
	Sens.	.253 (.061)	·	.501 (.053)	·	.650 (.032)	·
LMM	AUC	.633 (.047)	.80	.653 (.043)	.83	.663 (.040)	.90
	Sens.	.228 (.060)	.77	.460 (.052)	.74	.632 (.033)	.71
GEE	AUC	.631 (.047)	.82	.648 (.041)	.85	.656 (.040)	.91
	Sens.	.221 (.059)	.79	.445 (.051)	.90	.610 (.031)	.95
LR	AUC	.616 (.043)	.84	.642 (.043)	.87	.671 (.041)	.74
	Sens.	.210 (.058)	.80	.441 (.050)	92	.610 (.030)	.96

Table 4.4: AUC and sensitivity of JTIM and other models when readmission rates are varying.

4.5. Accuracy of the approximate transition probability

In 3.2, we derived an approximate formula that allows us to avoid numerical integration in calculating prediction probabilities. In this section we present a simulation study to examine

how well the approximation works in the computation of AUC and Brier scores.

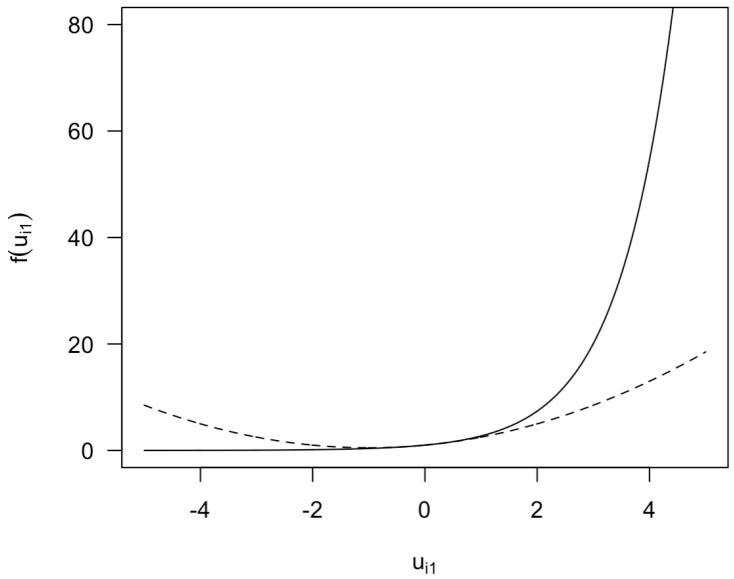

Figure 4.1: The solid line shows $e^{u_{i1}}$ and the dotted line shows $e^{u_{i1}} \approx 1 + u_{i1} + u_{i1}^2/2$

Figure 4.1 shows graphs of $e^{u_{i1}}$ and its Taylor expansion up to second order. We can see that the approximation works well when u_{i1} is between around -1.5 and 1.5. Because $u_{i1} \sim N(0, \sigma_1^2)$, we can anticipate that the approximation would work well if σ_{01} is small enough to generate a large majority of deviates in this range.

To examine this, we implemented a simulation with $\sigma_1^2 \in \{0.5, 1, 2\}$, where the normal distribution covers 99.75% of observations within $[-1.5, 1.5], [-3, 3]$, and $[-4.2, 4.2]$, respectively. According to Figure 4.1, we can expect the approximation to work properly with the first variance, but not with the others. The simulation is the same as the one described in Chapter 4.1, with 80% of the data used to fit the JTIM. Details of the simulation setting are as follows: $N = 500$, with prediction windows corresponding to readmission rate $10\%, 30\%, 50\%$. We ran $M = 1,000$ Monte-Carlo simulation for each setting.

To evaluate both discrimination and calibration of prediction, we used AUC and the Brier score which is the mean squared deviation between the predictive probability and the

true binary observation in S replications:

$$\text{BS} = \frac{1}{S}\sum_{s=1}^{S}(p_s - o_s)^2.$$

We present 1−BS so that a higher score means better prediction accuracy.

Table 4.5 describes the simulation results. As expected, there is no difference between the exact and approximate approach when $\sigma_1^2 = 0.5$. Interestingly, when $\sigma_1^2 = 1$, although it includes some area where the Taylor expansion deviates from σ_1^2, the mean difference of AUC is at most 0.002 and of 1-BS at most 0.005. The case of $\sigma_1^2 = 2$ produces some noticeable differences in 1-BS, implying that the approximation is starting to deteriorate at this level. Nevertheless, the differences are limited to 0.012 on the AUC scale and 0.019 on the 1-BS scale. On the whole, the approximation seems to be suitable for practical use, provided one gives due attention to the estimated value of the frailty variance.

	Readmission rate		10%	30%	50%
			Mean (se)	Mean (se)	Mean (se)
$\sigma_1^2 = 0.5$	Training set	AUC	.001 (.002)	0.0 (.002)	0.0 (.001)
		1-BS	.001 (.002)	.001 (.002)	.001 (.002)
	Test set	AUC	0.0 (.004)	0.0 (.003)	0.0 (.002)
		1-BS	.001 (.002)	.001 (.002)	.001 (.002)
$\sigma_1^2 = 1$	Training set	AUC	.002 (.003)	.002 (.002)	.001 (.002)
		1-BS	.003 (.003)	.005 (.003)	.003 (.002)
	Test set	AUC	.002 (.005)	.001 (.004)	.001 (.003)
		1-BS	.004 (.004)	.005 (.004)	.003 (.003)
$\sigma_1^2 = 2$	Training set	AUC	.009 (.008)	.011 (.011)	.012 (.013)
		1-BS	.018 (.016)	.019 (.014)	.016 (.010)
	Test set	AUC	.008 (.006)	.008 (.009)	.009 (.013)
		1-BS	.017 (.018)	.018 (.015)	.015 (.012)

Table 4.5: Difference between exact AUC (1-BS) and approximate AUC (1-BS)

CHAPTER 5

Applications

5.1. Exacerbation in patients with chronic bronchitis

We applied our models to analyze the data from a multicenter randomized trial involving patients with chronic bronchitis where the disease process is marked by the development and resolution of acute exacerbations of chronic bronchitis (AECB). The data is provided by Bayer Canada, Inc. During the study, each patient experienced alternating AECB and AECB-free periods. Our goal is to predict the probability of relapse of AECB within 2, 3, or 4 weeks after resolution of the previous episode.

To be eligible for the study, patients had to be 18 years or older, diagnosed with chronic bronchitis, able to maintain a daily diary, able to understand and complete detailed health status questionnaires, and currently experiencing an AECB. Patients who entered the study were randomized to receive either ciprofloxacin or standard care for a period of one year. They were required to visit the participating clinic when they perceived that a new exacerbation was beginning or an exacerbation was resolved. Patients were to be followed for 365 days, but early termination would occur if the subject either i) refused to complete the symptom diary, or ii) declined to return for further follow-up visits, or iii) died. There were 115 eligible patients randomized to take ciprofloxacin and 107 randomized to standard care. Patients were recruited from November 1993 to June 1994, with an average duration of follow-up of 357 days on ciprofloxacin and 350 days on standard care. The time scale is days. The average length of AECB was 15 days and AECB-free was 114 days.

In the simulation studies, we saw that med(r_i) is an important determinant of prediction performance. Because the (10, 25, 50, 75, 90) centiles of r_i in the AECB study are (0,0,2,3,4), we can expect the JTIM to perform well.

Predictor	**Description**	**Type**	**Baseline level**
Gender	Male and female	Nominal	Male
Length of AECB	Duration of previous AECB	Discrete	-
Symptoms	Preceding days of AECB symptom when a patient entered the study	Discrete	-
Severity	Severe chronic bronchitis or not	Nominal	Not severe
History	Years since diagnosis of chronic bronchitis	Discrete	-
Treatment	Ciprofloxacin or standard care	Nominal	Standard care
Season	Season when a patient entered AECB state Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec	Nominal	Oct-Dec

Table 5.1: Description of predictors for AECB data

Table 5.1 lists potential predictors in the AECB data. Some are common to both AECB and AECB-free events: Gender, severity, history, treatment, and season. To examine the possible effect of the time since diagnosis with chronic bronchitis, the history predictor indicates which of the following five-year intervals the patient history falls into: 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, 35–40, 40–45 years duration. We change this to an ordinal variable by choosing the center of the interval: 2.5, 7.5,...,42.5 years. Besides those, we included a symptoms predictor for the AECB state. For the AECB-free state, we included length of AECB.

For every AECB-free event, we predicted whether a relapse of AECB occurred within 2, 3, and 4 weeks after resolution of the previous episode. These windows correspond to relapse rates of roughly 10%, 15%, and 20%. We assume that each patient is censored instantly after entering an AECB-free event, so the last censored gap time is 1 day. We applied the four models from the simulation study with five-fold cross-validation to prevent overfitting bias. Again, we included the average annual number of past exacerbations as a predictor in

the models except JTIM.

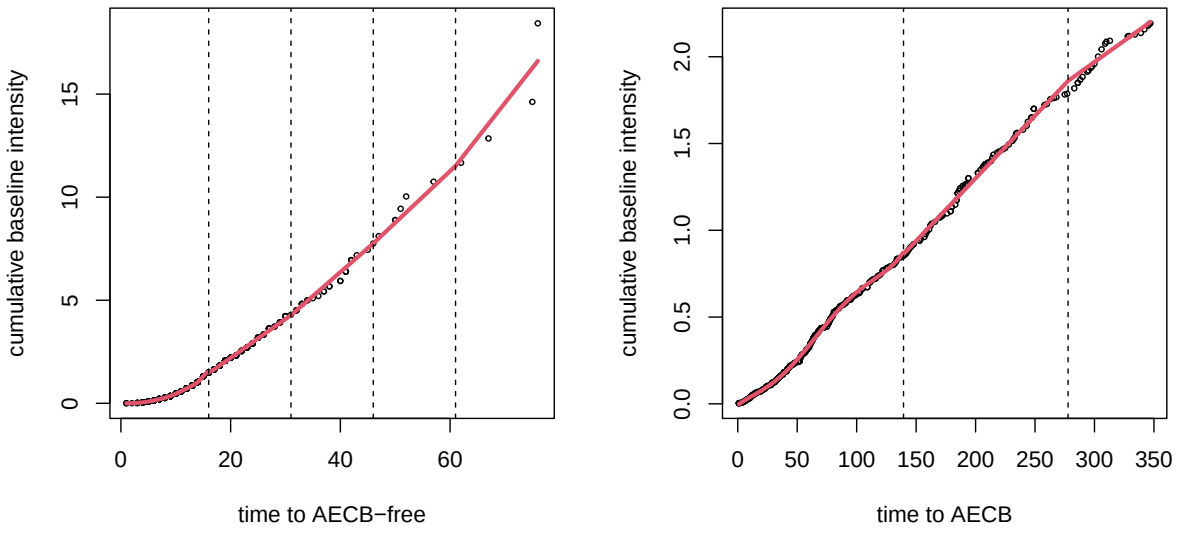

Figure 5.1: The estimated cumulative baseline transition intensity for time to AECB-free and to AECB.

Figure 5.1 shows the estimated B-spline cumulative baseline transition intensities for time to AECB-free and time to AECB. The solid line is an estimated B-spline regression and the dotted lines indicate knots. The intensity for time to AECB-free is far higher than for time to AECB, representing much longer time patients spent in the AECB-free state. The sparsity between points in the left plot indicates that there are few unique LoS values, whereas there are numerous LoD values in the right panel. For time to discharge, we can see a slight curve indicating that the baseline transition intensity varies across gap times. The estimated curve for time to AECB is more nearly linear, implying that the baseline intensity is roughly constant across gap time.

	AECB		AECB-free	
	Estimate	SE	Estimate	SE
Gender	0.162	0.121	-0.243	0.138
Length of AECB	—	—	**-0.018**	0.006
Symptoms	**-0.018**	0.009	—	—
Severity	**-0.305**	0.173	**0.587**	0.191
History	**-0.014**	0.006	**0.020**	0.007
Treatment	0.217	0.120	-0.078	0.136
Season(ref: Oct-Dec)				
Jan-Mar	-0.093	0.121	-0.236	0.155
Apr-Jun	0.140	0.135	**-0.340**	0.158
Jul-Sep	**-0.362**	0.151	0.001	0.182
SD of frailty	0.557		0.591	
Correlation	-0.56			

Table 5.2: Application to AECB data: estimated regression parameters (bolded when $P < 0.05$)

Table 5.2 shows the JTIM estimation results. Patients with longer preceding symptoms or longer bronchitis history are less likely to move to the AECB-free state. Also, patients with severe chronic bronchitis or longer history have a higher chance to relapse. Length of AECB has a statistically significant effect on reducing the chance of relapse. Compared to AECB occurring between October and December, onset of AECB between April and June reduces the chance of relapse, and onset between July and September decreases the chance of resolution.

The standard deviation of the frailty for time to AECB-free is 0.557, so the corresponding variance estimate is about 0.31, slightly smaller than 0.35, the variance estimate for time to AECB. The estimated correlation is -0.56; thus, a higher frailty of relapse is associated with a lower frailty for recovery.

Relapse window		2-week				3-week				4-week			
Models		JTIM	LMM	GEE	LR	JTIM	LMM	GEE	LR	JTIM	LMM	GEE	LR
Training set	AUC	.63 (.62)	.72	.68	.51	.66 (.66)	.70	.66	.53	.64 (.64)	.67	.61	.51
	Sens.	.11	.23	.08	.08	.19	.23	.09	.09	.25	.26	.11	.14
Test set	AUC	.64 (.64)	.54	.54	.58	.66 (.65)	.59	.58	.58	.64 (.64)	.59	.59	.54
	Sens.	.15	.11	.11	.03	.18	.18	.14	.13	.21	.23	.20	.15

(·): a result based on approximate transition probability

Table 5.3: Prediction results for AECB data. Each number is the average of a five-fold cross validation.

Prediction results appear in Table 5.3. LR is worst in both AUC and sensitivity. Compared to LMM and GEE, JTIM is lower by 9% on AUC for the training set, but it is 5–10% higher on the test set. JTIM shows nearly equal AUCs between training and test sets, suggesting it is less prone to overfitting than the other methods. When the relapse rate is low, JTIM outperforms LMM on sensitivity, but results flip when the relapse rate is high.

AUCs based on the approximate transition probability are close to the exact AUCs, reflecting the modest variance of the frailty of transition to the AECB-free state.

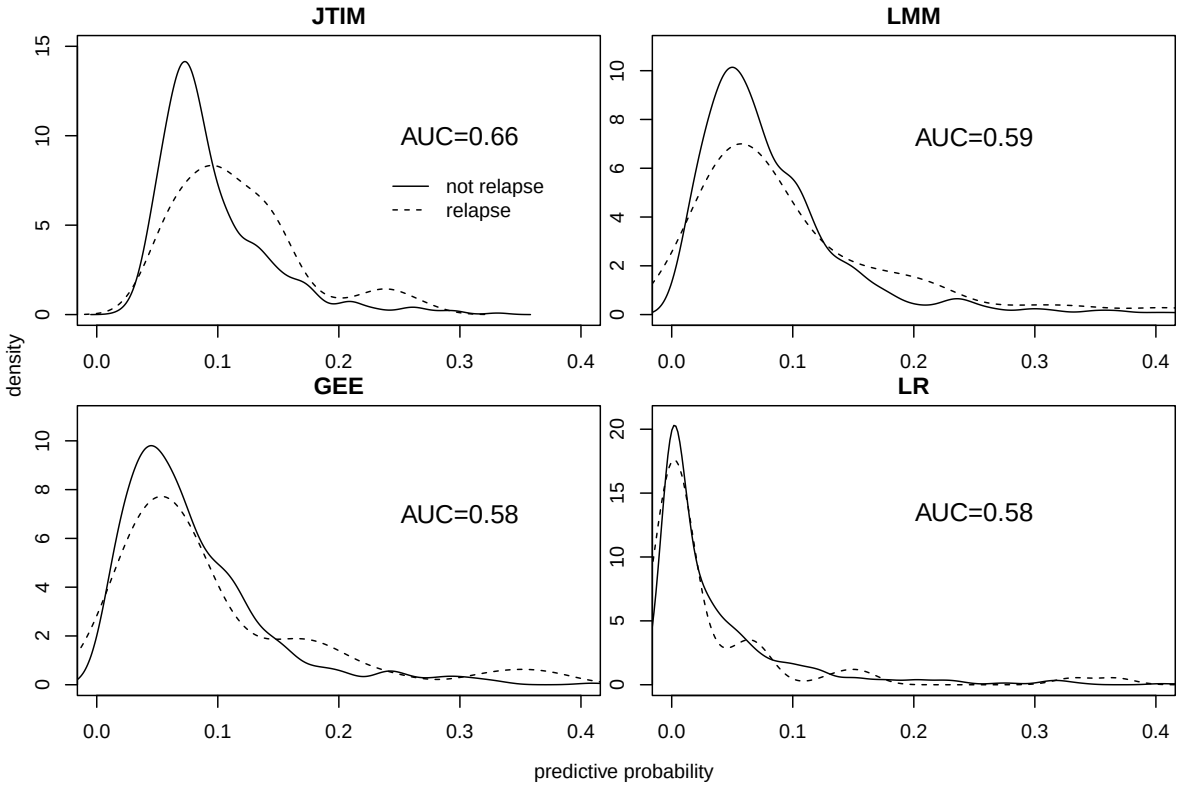

Figure 5.2: Smoothed histograms of predictive transition probabilities for the test set when the window of prediction is 3 weeks.

Figure 5.2 presents smoothed histograms of predictive transition probabilities for the test set when the window of prediction is three weeks. It is apparent that JTIM gives the most separation between the smoothed lines, reflecting its higher predictive power.

5.2. Readmission in patients with diabetes in Medicaid

Medicaid is a US federal and state program providing medical insurance for people with low incomes. The federal and state governments jointly fund it, with eligibility and coverage varying across states. It is the single largest source of health insurance in the US, covering nearly 76 million Americans under age 65 including children, pregnant women, low-income adults, and individuals with disabilities. Medicaid claims data comprise baseline information, institutional, medical, and pharmaceutical data for each patient encounter.

We sought to use Medicaid data to predict risk of 30-, 90-, and 180-day readmission of patients with diabetes hospitalized for any cause. We considered claims for patients with diabetes from plans in Georgia dated from January 1, 2016 to August 1, 2019. We identified diabetes patients from three sources: Admission ICD-10 codes in hospitalization claims; clinical diagnosis ICD-10 codes in physician and hospitalization claims; and prescriptions for diabetes drugs in pharmacy claims. We included in our cohort any patient who had at least one claim indicating diagnosis or treatment of non-gestational diabetes in any of those sources.

We collected all admission records from our cohort, defining an admission as any claim that was institutional, inpatient, and had a clear admission and discharge date. We limited these claims to begin from January 1, 2017, giving us at least one year of prior data for every admission record. Once we collected the records, we obtained the length of discharge (LoD) which is the difference between a subsequent admission date and a preceding discharge date. In this way we identified a cohort of 1,474 patients with diabetes. The time scale is days. The average length of stay (LoS) between patients was 6 days, and the average LoD was 243 days.

The (10, 25, 50, 75, 90) centiles of r_i are (0,0,0,1,2), implying that more than 50% of patients do not have a complete event pair. In fact, 60% have a pair of one uncensored admission claim and one censored discharge record. Because Table 4.2 in the simulation study shows that JTIM performs slightly better than the other models under these circumstances, we do not expect JTIM to dominate unless the variance of the frailty transiting to admission is high.

Covariate	Description	Type	Baseline level
Gender	Male and female	Nominal	Female
Age	Age of a patient	Discrete	-
Length of stay(LoS)	Duration of previous admission	Discrete	-
Diabetes type	Type I, II, other type	Nominal	Other type
Admission type	Emergency, urgent, elective, trauma, etc.	Nominal	Elective
Discharge type	Discharge home or an institution	Nominal	Home
Comorbidity	14 comorbidities: diagnosed or not	Nominal	Not diagnosed

Table 5.4: Covariates for the diabetes claims data

Table 5.4 lists the potential covariates for the diabetes claims data. As common covariates for both admission and discharge states, we used the following: Gender, age, diabetes type, admission type, and 14 comorbidities. We also included discharge type and LoS to predict readmission.

As in the AECB application, we applied the four models used in the simulation study with five-fold cross-validation. LMM, GEE, and LR used the average annual number of past admissions as a covariate. We predicted the probability of 30-, 90-, and 180-day readmission, corresponding to 16%, 33%, and 50% readmission rates, respectively. When predicted, we assume that each patient is censored immediately after entering discharge state, so the last censored gap time is one day.

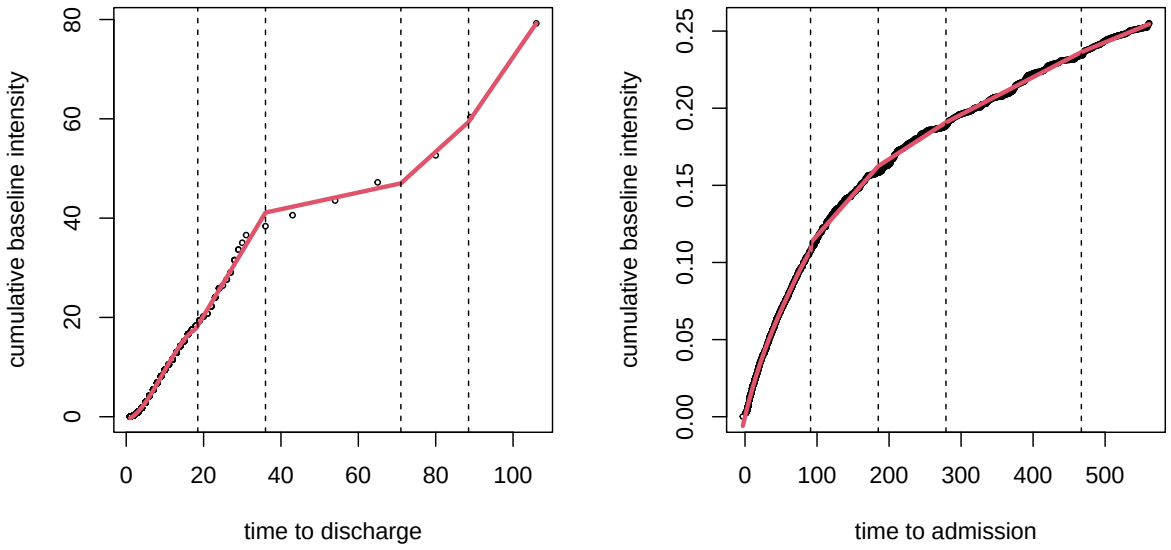

Figure 5.3: The estimated cumulative baseline transition intensity for time to discharge and admission.

Figure 5.3 presents estimated cumulative baseline transition intensities for time to discharge and admission. The intensity for time to discharge is far larger than for time to admission because hospitalizations are fairly short and sojourns outside of the hospital is long. Compared to LoD, LoS has smaller number of unique values; thus, the points in the left plot are sparse. The estimated cumulative intensity curves suggest that the baseline transition intensities are not constant for both.

	Admission to discharge		Discharge to admission	
	Estimate	SE	Estimate	SE
Gender	-0.034	0.088	**-0.328**	0.116
Age	**-0.016**	0.003	0.003	0.004
Length of stay	-	-	-0.004	0.007
Diabetes type (ref: other type)				
Type I	**0.061**	0.079	**0.669**	0.106
Type II	**0.144**	0.063	**0.216**	0.100
Admission type (ref: elective)				
Emergency	-0.027	0.072	**0.503**	0.113
Etc.	**-0.593**	0.294	**0.878**	0.381
Trauma	**-1.341**	0.406	**1.171**	0.578
Urgent	**-0.506**	0.085	0.244	0.134
Discharge type (ref: home)				
Institution	-	-	**0.616**	0.243
Comorbidity				
Renal failure	**-0.293**	0.079	0.144	0.097
Obesity	**0.119**	0.053	**-0.268**	0.088
Malignant neoplasms	0.044	0.166	0.246	0.204
Cerebrovascular	-0.027	0.117	0.060	0.146
Heart failure	-0.102	0.099	**0.224**	0.105
Cardiac arrhythmia	0.062	0.101	0.023	0.125
Hypertensive diseases	-0.122	0.070	0.039	0.096
Hemolytic anemia	-0.110	0.070	0.004	0.092
Cognitive function symptoms	**-0.228**	0.066	-0.122	0.088
Long term drug therapy	0.084	0.061	0.044	0.083
Need devices	-0.027	0.165	0.108	0.201
Nutritional metabolic disease	-0.061	0.077	-0.068	**0.100**
Personal history of several risk factors	**0.166**	0.066	0.052	0.086
Digestive	-0.006	0.061	**0.197**	0.082
SE of frailty	0.748		0.671	
Correlation	-0.43			

Table 5.5: Application to patients with diabetes admission data: estimated regression parameters (bolded when $P < 0.05$)

Table 5.5 shows estimates of the JTIM parameters. Among the baseline covariates, males are less likely to be readmitted than females, and older patients are less likely to be discharged than younger patients. LoS is not statistically significant for predicting readmission hazard. Both diabetes type and admission type show statistically significant effects on both transitions. Type I and II diabetes increase the chance to be discharged and re-hospitalized compared to other types of diabetes, implying that patients with type I and II diabetes experience hospitalization more frequently. Moreover, compared to elective admissions, the other types of admission decrease the possibility of discharge but increase the readmission hazard. Patients who are discharged to a skilled nursing facility or long-term care hospital have a higher risk of readmission than patients discharged home. Each past comorbidity contributes differently to possibility of discharge and readmission hazard. Interestingly, obesity is associated with an increased chance of discharge and a reduced chance of readmission.

The estimated variance of the frailty transiting to discharge is 0.56, and the variance of the frailty transiting to admission is 0.45. Referring to the simulation results in Table 4.3, we expect the JTIM to give modestly better AUC and sensitivity. The estimated correlation of -0.43 implies that a patient with a higher readmission frailty will have a reduced chance of early discharge.

Readmission window		30-day				90-day				180-day			
Models		JTIM	LMM	GEE	LR	JTIM	LMM	GEE	LR	JTIM	LMM	GEE	LR
Training set	AUC	.72 (.72)	.74	.73	.71	.75 (.75)	.76	.75	.74	.75 (.75)	.75	.75	.74
	Sens.	.40	.42	.40	.38	.55	.56	.55	.53	.64	.64	.64	.63
Test set	AUC	.70 (.69)	.68	.67	.67	.73 (.73)	.72	.70	.70	.73 (.73)	.73	.72	.72
	Sens.	.39	.35	.34	.34	.53	.52	.51	.48	.63	.61	.62	.59

(·): based on the approximate transition probability

Table 5.6: Prediction results for the Medicaid diabetes readmission data — averages of five-fold cross validation.

Table 5.6 presents AUC and sensitivity values for the diabetes data. On the whole, JTIM produces better or equal AUC and sensitivity compared to the other models, particularly

in the test set. To judge from the results in Table 4.2, the similar performance between models in 90-day and 180-day cases is caused by the absence of a history of admission in most patients ($\text{med}(r_i) = 0$). With the modest frailty variance of 0.45 for discharge state, AUCs based on the approximate transition probability differ by no more than 1% from exact values.

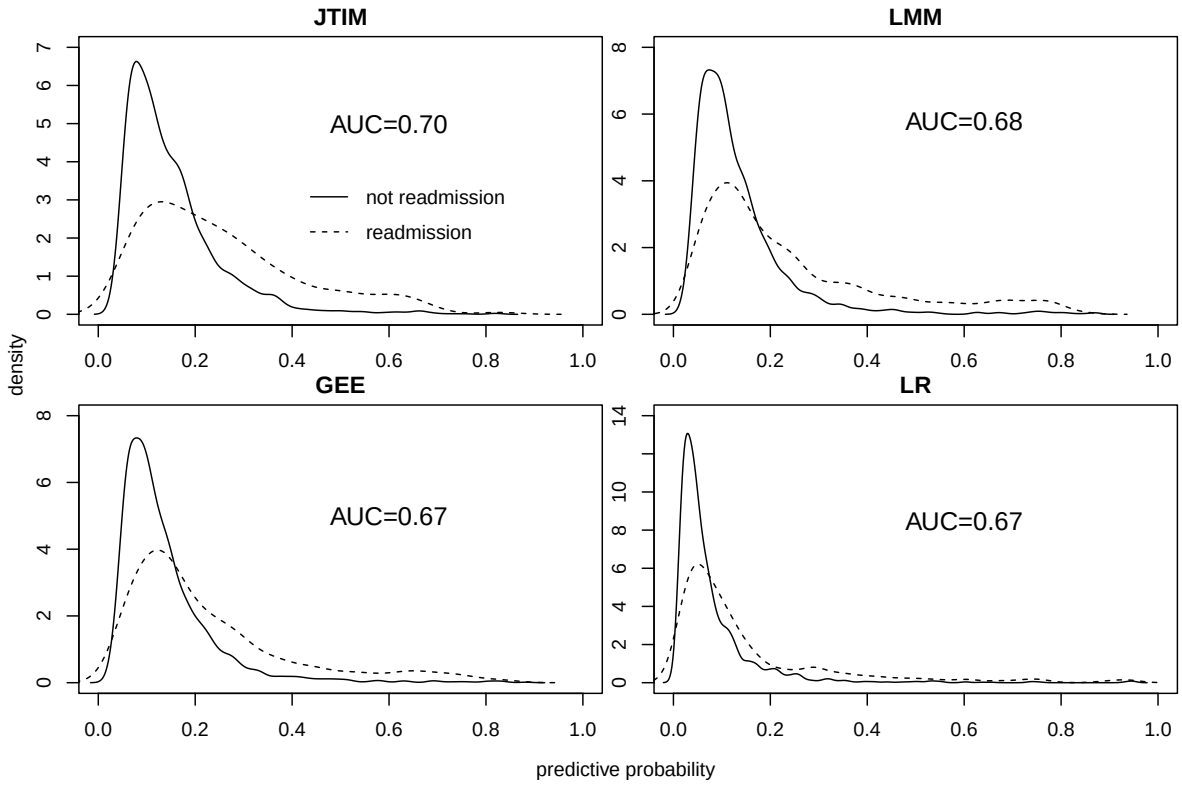

Figure 5.4: Smoothed histograms of predictive transition probabilities for the test set when the window of prediction is 30 days.

Figure 5.4 presents smoothed histograms of predictive transition probabilities for test set when the window of prediction is 30 days. Differences between the groups are modest in all models, reflecting their similar, relatively poor predictive power.

CHAPTER 6

Comparison Between Logistic Regression and Random forests for Longitudinal Claims Data with Binary Outcome

6.1. Motivation

Several studies have used electronic health record (EHR) or insurance claims data to predict 30-day readmission for patients with chronic diseases such as diabetes, heart failure (HF), and chronic obstructive pulmonary disease (COPD). Generally, this type of data has a longitudinal structure in that each patient has potentially a series of hospitalizations and readmissions during the window of follow-up. Yet few studies have sought to accommodate within-subject or other forms of clustering. For example, for data with patients having multiple hospitalization records, some studies use the first, last, or a randomly selected hospitalization record per patient to build a logistic regression (LR). Then, the model is applied to all observations for prediction. Recently, some authors have applied machine learning algorithms such as random forests (RF) or support vector machines (SVM) to predict readmission,[1] [10] [12] but these studies also have avoided the complexity of modeling clustering.

Various studies have compared machine learning algorithms with LR for predicting readmission. Reddy et al.[28] compared recurrent neural networks with long short-term memory (RNN-LSTM) with penalized LR for readmission of patients diagnosed with lupus. Lin et al.[22] addressed unplanned intensive care unit readmission prediction with RNN-LSTM, RF, SVM, and penalized LR. For HF readmission, Frizzell et al.[10] compared the prediction performance between a tree-augmented naive Bayesian network, RF, gradient-boosted,

and LR using the Get With the Guidelines Heart Failure (GWTG-HF) registry linked with Medicare inpatient data. Huang et al.[15] and Shin et al.[35] systematically reviewed the literature comparing machine learning algorithms including RF, neural networks, and SVM to conventional statistical models such as LR and Cox regression. Min et al.[26] compared the performance of different machine learning models for predicting the hospital readmission risk of COPD patients.

Most of these articles do not address the performance of methods applied to longitudinal observations. When they fitted LR, some treated all the repeated measurments as independent, ignoring the correlation within subjects and other clusters. Some comparisons used a single index admission per patient, making comparisons under longitudinal EHR or claims data still unclear. Frizzell et al.[10] was the only study to at least make the comparison valid, by choosing the first hospitalization within each patient and using it as an index admission to build models.

Finally, studies comparing out-of-sample prediction performance are lacking. A model derived from Texas EHR data, for example, likely predicts well for Texas but not quite as well for other states, although still well enough to be practically useful. We conjecture that LR models capture the broad dependence of binary outcomes on predictors in a way that transcends differences between data collection units such as states, whereas machine learning techniques such as RF give predictions that are exquisitely tuned to one state but perhaps less useful for others. This question, which we have not seen discussed in the literature, is amenable to empirical investigation in our Medicaid data, which includes claims from four US states.

In this chapter, we compare the prediction performance of LR and RF for 30-day readmission with longitudinal claims data. We define RF as Leo Breiman's original version of random forests [3]. Using simulated longitudinal claims data, first we randomly choose a record from each patient in the training set and fit the two models, applying them to the training, test, and external sets to compute AUC and sensitivity. We validate this comparison through an empirical study using Medicaid claims data, comparing in- and out-of-sample

prediction for data from four states.

6.2. Brief Overview

For observation i, let y_i be the outcome and $\boldsymbol{X}_i$ denote a covariate vector. Then, the distribution of y_i given $\boldsymbol{X}_i$ follows the binomial distribution

$$Y_i \sim \text{Bin}(1, \pi_i),$$

where $\pi_i = \Pr(Y_i = 1)$. The logistic model is defined as

$$\pi_i = \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)}, \qquad i = 1, ..., n$$

where $\boldsymbol{\beta}$ is a vector of coefficients. Alternatively, one can write

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \boldsymbol{\beta}^T \boldsymbol{X}_i.$$

Because LR lies within the class of generalized linear models, it is straightforward to estimate and validate. The fact that LR coefficients represent odds ratios renders its results readily interpretable in any context, especially compared to ensemble-based learning algorithms that combine various predictions in a complex way. LR can fail, however, if the logit success probability depends on predictors in a nonlinear way or with multiple interactions.

RF relies on its predecessor method *classification and regression trees (CART)*, a supervised machine learning algorithm that describes the conditional distribution of response Y given $\boldsymbol{X}$. Note that since our interest is classification of binary outcome (30-day readmission) rather than regression for continuous variable, we are focusing on the classification tree here. CART uses a greedy recursive binary splitting algorithm to partition the feature space into rectangles (regions), defining an estimated class for each one. It starts from the root node and divides the input space into two branches at each decision node, repeating this

process until all branches reach terminal nodes. Figure 6.1 illustrates a classification tree

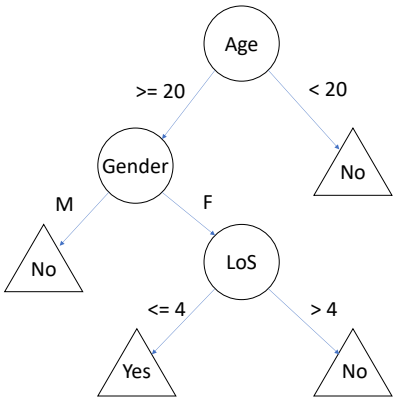

Figure 6.1: Example of classification tree for binary outcome

for a binary outcome. The root node is age, and decision nodes are gender and length of stay. Each feature is partitioned once; at the end there are four terminal nodes.

To identify the splitting variable and level, it is typical to optimize the Gini index or cross-entropy, which measure homogeneity of the classified sample. For two classes, those measures equal 0 when the classes are perfectly homogeneous, or 1 when class labels are equally divided.

If the size of the tree is too large, then the model might overfit the data; whereas if the size is too small, the model can miss important structure. To determine the optimal size of a tree, a general strategy is cost-complexity pruning. For example, if the cost of adding another variable to the tree from the current node exceeds the value of the complexity parameter of the current tree, then tree-building ends.

Suppose that there are N subjects, regions in the model are already known, and they are indexed by d. Then, the problem is to estimate a class for each node. In a node d representing region R_d with n_d observations, we denote the proportion of class k observations in node d as

$$\hat{p}_{dk} = \frac{1}{n_d} \sum_{\boldsymbol{X}_i \in R_d} I(y_i = k), \qquad i = 1, ..., N$$

Then we classify the observations in node d to class

$$k(d) = \arg\max_k \hat{p}_{dk}$$

which is the majority class in node d. For the binary problem, k has two classes.

CART is a flexible method that easily handles nonlinear and non-smooth relationships. Moreover, the tree plots (see Figure 6.1) provide an intuitive interpretation for its binary structure. A significant drawback of CART is overfitting, which results in high model variance, or tree instability. Because the method is hierarchical, a small change in the data can substantially alter the structure of the resulting tree. A change in a feature in a root node may result in a tree with completely different decision and terminal nodes.

The *random forest (RF)* addresses these concerns to some extent by the use of *bagging*, also called bootstrap aggregating [2]. This involves generating B bootstrap samples, fitting a tree to each, and averaging the B tree predictors for responses or calculating the majority vote of the predictors for classification. Because bagging uses bootstrap samples from the same original data, bagging trees are positively correlated, which increases model variance. RF corrects this problem by constructing de-correlated bagging trees. Growing an RF tree involves repeating the following steps:

1. Instead of using all p variables in the bootstrapped data, select $m \ll p$ variables at random.
2. Pick the best variable and its split-point among the m.
3. Split the node into two following nodes.

Other than that, the rest of the process is the same as bagging.

As well as reducing the model variance and preventing overfitting, the RF can use the out-of-bag samples left over from the bootstrap to measure the importance of each variable. Even so, a main weakness of RF is its lack of reproducibility, as trees in the forests are randomly built. And like other ensemble learning methods, the model has no obvious interpretation.

6.3. Simulations

To compare LR and RF for predicting 30-day readmission, we created simulated claims data following the description in 4.1. $N = 1000$ patients experienced recurrent admission and discharge events over a span of 3 years. We set discharge/readmission frailties to follow a bivariate normal distribution with both means 0 and both variances 0.5, and with correlation $\rho = -0.5$. The resulting simulated data has 10 days and 205 days as the average of LoS and LoD, respectively, and the $(10, 25, 50, 75, 90)$ centiles of r_i over patients are $(0, 1, 3, 4, 6)$; the readmission rate is around 22%. As before, we used 80% of the data as a training set, with the rest reserved as a test set.

We used both relevant and noise covariates, reflecting the situation in practice where an analyst may unwittingly include in a prediction model one or more irrelevant variables. Relevant variables included gender and age with true coefficients the same as in 4.1: $\boldsymbol{\beta}_0 = (-0.05, -0.01)^T$ for admission and $\boldsymbol{\beta}_1 = (-0.5, 0.001)^T$ for discharge. We also included length of stay of the previous admission and annual number of past admissions as relevant covariates. We included 5 pure noise variables generated independently from $N(0, 1)$. In summary, there are 4 relevant variables — gender, age, length of stay, and the annual number of past admissions — and 5 continuous noise variables.

In addition to the test set, we applied prediction models to external data which has similar properties with the data but is independently generated with different coefficients for gender and age: $\boldsymbol{\beta}_0 = (-0.02, -0.005)^T$ for admission and $\boldsymbol{\beta}_1 = (-0.2, 0.01)^T$ for discharge. Also, the mean of frailties is still 0 and variances are negatively correlated $\rho = -0.5$ while we changed σ_0^2 and σ_1^2 to 0.25. This out-of-sample data has 6 days and 240 days as the average of LoS and LoD, respectively, and the $(10, 25, 50, 75, 90)$ centiles of r_i are $(0, 1, 2, 3, 4)$, implying fewer past records per patient. The readmission rate is round 13%.

The simulation proceeds as follows:

1. Generate the data and split it into training and test sets.

2. Generate the external data.

3. Using the training set, randomly choose one index admission per patient.

    3-1. Fit the LR.

    3-2. With the fitted model, obtain predictive transition probabilities for 30-day readmission for the training set, test set, and external data.

    3-3. With the predictive transition probabilities, calculate AUC and sensitivity (as in 4.2) for the training set, test set, and external data.

    3-4. Repeat 3-1 through 3-3 with RF. Here, predictive probabilities are estimated by the fraction of times a target index admission is classified into the 30-day readmission class between trees in forests.

4. Repeat steps 1–3 10,000 times and average the prediction measures across replications.

Before fitting RF in step 3-1, we first tuned hyperparameters in RF to prevent overfitting. There are three hyperparameters: The number of features randomly sampled as candidates at each split, the number of trees to grow, and the maximum number of terminal nodes that trees in the forest can have. We considered 2–6 out of nine features for the first one; 50, 100, 500, 1000 trees for the second one; and 5–15 terminal nodes for the third one. We identified the optimal combination by a grid search. For each combination, we applied 5-fold cross-validation and measured the fraction of correct classifications, denoted the *accuracy*. Choosing the combination with the highest accuracy, we applied these parameters when fitting the RF in step 3-1.

Model		LR	RF	
		Mean (se)	Mean (se)	Win rate
Training set	AUC	.60 (.018)	.59 (.012)	.68
	Sens.	.33 (.006)	.28 (.007)	.80
Test set	AUC	.60 (.021)	.56 (.020)	.93
	Sens.	.32 (.014)	.24 (.013)	.95
External data	AUC	.58 (.018)	.55 (.012)	.90
	Sens.	.20 (.004)	.14 (.005)	.95

Table 6.1: Prediction performance comparison between logistic regression and random forests for longitudinal claims data

Table 6.1 presents the simulation results. Here the win rate which is the fraction of 10,000 Monte-Carlo simulations in which LR has superior performance to RF, calculated separately for AUC and sensitivity. Compared to RF, LR shows higher prediction performance for both AUC and sensitivity with the training, test, and external sets. Although the performance gap between LR and RF with the training set is modest, it increases to 0.04 with the test and external sets. The win rate shows a similar trend. If we compare the performance by each set, there is no evidence of overfitting in either model, but LR predicts better for the test set and is more robust with external data.

6.4. Empirical study: Readmission in Medicaid patients with heart failure

We conducted an empirical study in Medicaid patients admitted for heart failure (HF). Considering their HF admission records as index admissions, our goal was to predict all-cause 30-day readmissions.

We first considered Medicaid claims for patients with HF in 4 states — Colorado, Kentucky, Nevada, and Ohio — dated from January 1, 2016 to August 1, 2019. We identified HF patients from admission ICD-10 codes in hospitalization claims and clinical diagnosis ICD-10 codes in physician and hospitalization claims. Next, we collected their HF index

admissions and all subsequent admissions. We limited the records to include those dated January 1, 2017 or later, giving us one year of prior data for every index admission. Both index admissions and subsequent admissions were required to be i) institutional claims and ii) inpatient hospital claims, and iii) have clear admission and discharge dates. To be an HF index admissions, an admission record must include HF record(s) in either admission diagnosis or clinical diagnosis. Subsequent admission claims — i.e., potential readmissions — did not require an HF diagnosis. For each HF index admission record, we checked whether 30-day readmission occurred by calculating LoD, which is the difference in days between the admission date of the subsequent admission and the discharge date of the index admission.

State	CO	KY	NV	OH
Patients	1921	788	1925	1467
Age quantiles (25-50-75th)	47-56-62	49-57-63	47-55-60	49-56-61
% of female	41.6	53.3	38.8	54.1
Avg. LoS (days)	6.3	6.5	6.5	7.6
Avg. LoD (days)	216	192	179	245
Readmission rate (%)	32.4	27.2	33.7	21.6
med(r_i)	0	0	0	0

Table 6.2: Descriptive statistics by state (CO = Colorado, KY = Kentucky, NV = Nevada, OH = Ohio)

Table 6.2 presents descriptive statistics on index admissions by state. There is substantial heterogeneity across states in the number of patients, percentage female, average LoD, and readmission rates. All states show zero in the median number of past admissions, implying that 50% of patients or more have no past admission record. In fact, in any state, about 70% of patients have only one uncensored admission claim with a censored discharge record.

Variable	Description
Age	Discrete, 0-65
Gender	Male and Female
Discharge status	To home or an institution
Acuity	Emergent admission or not
Length of stay	Discrete, ≥ 1 day
Avg. annual # of past admissions	Continuous, ≥ 0
Avg. annual # of past ER visits	Continuous, ≥ 0
9 comorbidities (binary)	Renal failure
	Obesity
	Sepsis
	Respiratory failure
	Influenza pneumonia
	Hemolytic anemia
	Abnormal findings on blood
	Chest & throat & upper abdomen
	Personal history of several risk factors

Table 6.3: List of covariates. (ER: emergency room)

Table 6.3 lists the variables used in the predictive modeling. Most are equivalent to those in 5.2. Instead of using admission type with five levels, we re-grouped into two levels — emergency or not; we henceforth call this variable *acuity*. We moreover added the average annual number of past emergency room (ER) visits, which is likely to be related to frequent readmissions.

Analysis steps for the empirical study are as follows:

1. Choose one state and perform a 5-fold cross validation:

    1-1. Choose one of the five fractions as the test set and designate the rest as the training set.

    1-2. In the training set, choose the first index admission per patient.

    1-3. Fit the LR with the data from step 1-2.

1-4. Using the fitted model, obtain predictive transition probabilities for 30-day readmission for the training set, test set, and external sets (the data from the other three states).

1-5. With the predictive transition probabilities, calculate AUC and sensitivity for the prediction sets.

1-6. Repeat 1-3 to 1-5 with RF. Here, predictive probabilities are the fraction of times a target index admission is classified as a 30-day readmission class between trees in forests.

1-7. Repeat 1-1 to 1-6 for the other 4 test sets and average each measure for each training set, test set, and the three external sets by cross-validation.

2. Repeat step 1 for the other three states.

State		CO		KY		NV		OH	
Models		LR	RF	LR	RF	LR	RF	LR	RF
Training set	AUC	.84	.82	.75	.77	.73	.74	.66	.67
	Sens.	.65	.62	.51	.56	.54	.57	.38	.42
Test set	AUC	.75	.74	.75	.73	.75	.73	.64	.63
	Sens.	.58	.58	.53	.50	.60	.60	.36	.36

Table 6.4: Prediction results for training and test sets by state.

Results appear in Table 6.4. Overall, RF shows better performance with training set while LR predicts better with test set. However, in terms of AUC and sensitivity of test set, there are no significant performance gaps for both AUC and sensitivity between LR and RF except Kentucky. Hence, we can conclude that both models work analogously in most of states.

External data from		Model under CO		KY		NV		OH	
		LR	RF	LR	RF	LR	RF	LR	RF
CO	AUC			.80	.78	.80	.80	.79	.76
	Sens.			.59	.57	.65	.64	.53	.54
KY	AUC	.71	.69			.72	.72	.71	.62
	Sens.	.54	.51			.57	.54	.41	.34
NV	AUC	.71	.68	.71	.69			.70	.65
	Sens.	.53	.50	.45	.45			.39	.36
OH	AUC	.64	.62	.64	.61	.64	.64		
	Sens.	.48	.45	.41	.39	.48	.48		

Table 6.5: Prediction results for external data by state.

Table 6.5 presents prediction results for the external (out-of-sample) data, with the column indicating the state used to construct the model and the row indicating the state in which the prediction model was assessed. Notably, in almost every combination, LR gives AUC and sensitivity at least as good as RF, and often much better.

Despite the lack of reproducibility and interpretability, a recent trend in the literature has been applying RF with the expectation of higher prediction performance of ensemble-based algorithm. Nonetheless, our simulations and empirical study suggest that it is hard to say that RF predicts better than LR for 30-day readmission with longitudinal claims data. Rather, our results recommend LR showing higher robustness against test and external sets.

CHAPTER 7

Discussion

In Chapters 1–5, we proposed dynamic prediction for alternating recurrent events under the joint transition intensity model (JTIM). This model flexibly accommodates covariates, random effects, and within-subject correlation of repeated events. The method of Wang et al. provides fast simultaneous estimation of model parameters *via* the penalized partial likelihood. Because their approach sacrifices estimation of the baseline transition intensity function, we apply the Breslow estimator with B-spline regression. This allows us to account for potential nonlinear or nonsmooth effects of event duration on the baseline intensity function.

We propose a method to compute the predictive transition probability from one event to the other by integrating out the unknown transition frailty of new subject with respect to its current posterior distribution. Our approach uses all available information in the likelihood. We avoid numerical integration through application of a Taylor expansion.

Simulation studies show that our approach generally outperforms other methods that are based on logistic regression and its extensions to correlated data. Further, we have verified that the transition probability approximation is accurate even with large frailty variance. Two applications of our method to real data demonstrate its practicality in clinical prediction contexts.

The value of our work is in the modeling and prediction of alternating recurrent events, which occur in a broad range of applications in biomedicine, engineering, and finance. In the context of hospital readmission, our work can serve as a foundation for pre-/post-discharge management tools aimed at preventing avoidable readmissions.

We have formulated the model as a sequence of alternating events nested within a subject. In readmission, exacerbation of chronic disease, and other medical applications, there may also be correlation at the level of the provider, hospital, or other units. One can address this in statistical models by including additional frailty terms, for example at the hospital level. Let the vector of hospital transition frailties $\boldsymbol{v}_l = (v_{l0}, v_{l1})^T$ follow a bivariate normal distribution. Also, presume that it is independent of patient transition frailties i.e. $\boldsymbol{u}_i \perp\!\!\!\perp \boldsymbol{v}_l, \forall(i,l)$. The maximum value of the hospital index l should be less than or equal to the number of patients n, as presumably some patients use the same hospital. One can then express the marginal transition probability as

$$\begin{aligned}
&\Pr(W^*_{i1(r_i+1)} \leq C^w_{i1(r_i+1)} + \tau | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{v}_l}) \\
&= \int_{v_{l1}} \int_{u_{i1}} \Pr(W^*_{i1(r_i+1)} \leq C^w_{i1(r_i+1)} + \tau | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{v}_l}, u_{i1}, v_{l1}) \\
&\quad \times f_{u_{i1}v_{l1}}(u_{i1}, v_{l1} | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{v}_l}) du_{i1} dv_{l1},
\end{aligned}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$ and $\boldsymbol{\Sigma}_{\boldsymbol{v}_l}$ are the variance-covariance matrix of the patient and hospital frailties, respectively. The joint posterior density of u_{i1} and v_{l1}, the second integrand, is

$$\begin{aligned}
&f_{u_{i1}v_{l1}}(u_{i1}, v_{l1} | W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)}, H_i, \boldsymbol{Z}_i, \boldsymbol{\xi}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{v}_l}) \\
&= \frac{f_{C^w_{i1(r_i+1)}}(W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)} | u_{i1}, v_{l1}, \boldsymbol{Z}_i, \boldsymbol{\xi}) f_{H_i}(H_i | u_{i1}, v_{l1}, \boldsymbol{Z}_i, \boldsymbol{\xi}) f_{u_{i1}}(u_{i1} | \boldsymbol{\Sigma}_{\boldsymbol{u}_i}) f_{v_{l1}}(v_{l1} | \boldsymbol{\Sigma}_{\boldsymbol{v}_l})}{\int_{v_{l1}} \int_{u_{i1}} f_{C^w_{i1(r_i+1)}}(W^*_{i1(r_i+1)} \geq C^w_{i1(r_i+1)} | u_{i1}, v_{l1}, \boldsymbol{Z}_i, \boldsymbol{\xi}) f_{H_i}(H_i | u_{i1}, v_{l1}, \boldsymbol{Z}_i, \boldsymbol{\xi}) f_{u_{i1}}(u_{i1} | \boldsymbol{\Sigma}_{\boldsymbol{u}_i}) f_{v_{l1}}(v_{l1} | \boldsymbol{\Sigma}_{\boldsymbol{v}_l}) du_{i1} dv_{l1}}.
\end{aligned}$$

We can derive the rest in the same way, as in 3.1.

These formulas apply when each patient uses only one hospital, which is not necessarily the case as patients may not always be admitted to the same hospital when ill. Such arbitrary crossing of patients with hospitals renders the problem more difficult computationally.

In Chapter 6, we examined the prediction performance of LR and RF for longitudinal claims data. By selecting the first observation or randomly choosing one observation per patient, we built prediction models that are valid although possibly inefficient. Including

training and test sets, we predicted external data as well to examine flexibility and robustness. Through simulations and empirical study, we verified that the LR model generally performs as well as or better than RF for the test set and is more robust in out-of-sample predictions.

Ensemble learning consists of two steps: Constructing a set of models from different training sets, and then combining predictors from them to produce one output per response. Ensemble learning can reduce risk compared to a model chosen and built by one-time data-splitting like cross-validation [8]. When a feature space is too large for training set, prediction accuracy from such a randomly chosen model would be doubtful. From a statistical point of view, this is a way to decrease model variance. Such models may be difficult to interpret, however. Our examination suggests that there is no reason to prefer RF over LR to predict 30-day readmission for longitudinal claims data.

Kirasich et al. [17] developed an analytical tool to compare LR and RF with simulated data. The data generated is not longitudinal but encompasses several options: The number of observations; the number of relevant and noise variables; the presence of continuous and categorical variables; the regression coefficients; and the variability of errors. We can refer to this tool to do further comparisons of prediction models. For instance, RF can handle large numbers of variables, so simulations in the large p setting may be useful. Further, simulations of longitudinal claims data with different properties such as divergent $\text{med}(r_i)$ or readmission rate will be interesting. For example, applying simulated data with a greater $\text{med}(r_i)$ can demonstrate performance in this setting. Also, with a higher readmission rate, we can see how well a trained model captures the information of readmission. Finally, we recommend more studies with various forms of heterogeneity in the external set.

APPENDIX A

A.1. Approximation

Recall that by Taylor expansion, $e^{u_{i1}} = 1 + u_{i1} + \frac{u_{i1}^2}{2!} + o(u_{i1}^2)$ as $u_{i1} \to 0$. Letting $\eta(W_{ijk}) = \Lambda_{01}(W_{ijk}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k})$, then

$$\begin{aligned} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) &= \exp\Big(- \Lambda_{01}(C^w_{i1(r_i+1)}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k} + u_{i1})\Big) \\ &= \exp\Big(- \eta(C^w_{i1(r_i+1)}) e^{u_{i1}}\Big) \\ &\approx \exp\Big(- \eta(C^w_{i1(r_i+1)})(1 + u_{i1} + u_{i1}^2/2)\Big) \end{aligned}$$

and

$$\begin{aligned} \prod_{k=1}^{r_i}\Big(- S'_{i1k}(W_{i1k})\Big)^{\delta_{i1k}} &= \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k} + u_{i1}) \exp\Big(- \Lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k} + u_{i1})\Big) \\ &= \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}) e^{u_{i1}} \exp\Big(- \eta(W_{i1k}) e^{u_{i1}}\Big) \\ &= \exp\Big(r_i u_{i1} - e^{u_{i1}} \sum_{k=1}^{r_i} \eta(W_{i1k})\Big) \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}) \\ &\approx \exp\Big(r_i u_{i1} - (1 + u_{i1} + u_{i1}^2/2) \textstyle\sum_{k=1}^{r_i} \eta(W_{i1k})\Big) \displaystyle\prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}). \end{aligned}$$

First, we approximate the denominator in 3.5 as follows:

$$\int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k}) \Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$\approx \int_{u_{i1}} \exp\Big(- \eta(C^w_{i1(r_i+1)})(1 + u_{i1} + u^2_{i1}/2) \Big) \exp\Big(r_i u_{i1} - (1 + u_{i1} + u^2_{i1}/2) \sum_{k=1}^{r_i} \eta(W_{i1k}) \Big)$$

$$\times \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}) f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}.$$

In 2.1, we define the observed gap time $W_{ijk} = \min(W^*_{ijk}, C^w_{ijk})$. Since the last gap time of type 1 event is censored, $W_{i1(r_i+1)} = C^w_{i1(r_i+1)}$ and this leads to $\sum_{k=1}^{r_i} \eta(W_{i1k}) + \eta(C^w_{i1(r_i+1)}) = \sum_{k=1}^{r_i+1} \eta(W_{i1k})$. Therefore,

$$\int_{u_{i1}} \exp\Big(- \eta(C^w_{i1(r_i+1)})(1 + u_{i1} + u^2_{i1}/2) \Big) \exp\Big(r_i u_{i1} - (1 + u_{i1} + u^2_{i1}/2) \sum_{k=1}^{r_i} \eta(W_{i1k}) \Big)$$

$$\times \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}) f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$= \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}) \int_{u_{i1}} \exp\Big(r_i u_{i1} - (1 + u_{i1} + u^2_{i1}/2) \sum_{k=1}^{r_i+1} \eta(W_{i1k}) \Big) f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$= \exp\Big(- \sum_{k=1}^{r_i+1} \eta(W_{i1k}) \Big) \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k})$$

$$\times \int_{u_{i1}} \exp\Big(r_i u_{i1} - (u_{i1} + u^2_{i1}/2) \sum_{k=1}^{r_i+1} \eta(W_{i1k}) \Big) f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}.$$

For simplicity, let $c_0 = \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k})$ and $c_1 = \sum_{k=1}^{r_i+1} \eta(W_{i1k})$ where both are not the function of u_{i1}. Then,

$$\exp\Big(-\sum_{k=1}^{r_i+1}\eta(W_{i1k})\Big)\prod_{k=1}^{r_i}\lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k})$$

$$\times\int_{u_{i1}}\exp\Big(r_iu_{i1}-(u_{i1}+u_{i1}^2/2)\sum_{k=1}^{r_i+1}\eta(W_{i1k})\Big)f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}$$

$$=e^{-c_1}c_0\int_{u_{i1}}\exp\Big(r_iu_{i1}-c_1(u_{i1}+u_{i1}^2/2)\Big)f_{u_{i1}}(u_{i1}|\sigma_1)du_{i1}$$

$$=e^{-c_1}c_0\int_{u_{i1}}\exp\Big(r_iu_{i1}-c_1(u_{i1}+u_{i1}^2/2)\Big)\frac{1}{\sqrt{2\pi}\sigma_1}\exp\Big(-\frac{u_{i1}^2}{2\sigma_1^2}\Big)du_{i1}$$

$$=\frac{e^{-c_1}c_0}{\sqrt{2\pi}\sigma_1}\int_{u_{i1}}\exp\left(r_iu_{i1}-c_1(u_{i1}+u_{i1}^2/2)-\frac{u_{i1}^2}{2\sigma_1^2}\right)du_{i1}$$

$$=\frac{e^{-c_1}c_0}{\sqrt{2\pi}\sigma_1}\int_{u_{i1}}\exp\left(-\frac{u_{i1}^2(1+c_1\sigma_1^2)-2u_{i1}(r_i\sigma_1^2-c_1\sigma_1^2)}{2\sigma_1^2}\right)du_{i1}$$

$$=\frac{e^{-c_1}c_0}{\sqrt{2\pi}\sigma_1}\int_{u_{i1}}\exp\left(-\frac{u_{i1}^2-2u_{i1}\dfrac{r_i\sigma_1^2-c_1\sigma_1^2}{1+c_1\sigma_1^2}+\left(\dfrac{r_i\sigma_1^2-c_1\sigma_1^2}{1+c_1\sigma_1^2}\right)^2-\left(\dfrac{r_i\sigma_1^2-c_1\sigma_1^2}{1+c_1\sigma_1^2}\right)^2}{2\sigma_1^2/(1+c_1\sigma_1^2)}\right)du_{i1}$$

$$=\frac{e^{-c_1}c_0}{\sqrt{2\pi}\sigma_1}\int_{u_{i1}}\exp\left(-\frac{\Big(u_{i1}-\dfrac{r_i\sigma_1^2-c_1\sigma_1^2}{1+c_1\sigma_1^2}\Big)^2}{2\sigma_1^2/(1+c_1\sigma_1^2)}+\frac{(r_i\sigma_1^2-c_1\sigma_1^2)^2(1+c_1\sigma_1^2)}{2\sigma_1^2(1+c_1\sigma_1^2)^2}\right)du_{i1}$$

$$=\exp\left(\frac{(r_i\sigma_1^2-c_1\sigma_1^2)^2(1+c_1\sigma_1^2)}{2\sigma_1^2(1+c_1\sigma_1^2)^2}\right)\frac{e^{-c_1}c_0}{\sqrt{2\pi}\sigma_1}\int_{u_{i1}}\exp\left(-\frac{\Big(u_{i1}-\dfrac{r_i\sigma_1^2-c_1\sigma_1^2}{1+c_1\sigma_1^2}\Big)^2}{2\sigma_1^2/(1+c_1\sigma_1^2)}\right)du_{i1}.$$

Since $\int_{u_{i1}} \exp\left(-\frac{\left(u_{i1} - \frac{r_i\sigma_1^2 - c_1\sigma_1^2}{1+c_1\sigma_1^2}\right)^2}{2\sigma_1^2/(1+c_1\sigma_1^2)} \right) du_{i1} = \frac{\sqrt{2\pi}\sigma_1}{\sqrt{1+c_1\sigma_1^2}}$,

$$\exp\left(\frac{(r_i\sigma_1^2 - c_1\sigma_1^2)^2(1+c_1\sigma_1^2)}{2\sigma_1^2(1+c_1\sigma_1^2)^2}\right)\frac{e^{-c_1}c_0}{\sqrt{2\pi}\sigma_1}\int_{u_{i1}} \exp\left(-\frac{\left(u_{i1} - \frac{r_i\sigma_1^2 - c_1\sigma_1^2}{1+c_1\sigma_1^2}\right)^2}{2\sigma_1^2/(1+c_1\sigma_1^2)} \right) du_{i1}$$

$$= \exp\left(\frac{(r_i\sigma_1^2 - c_1\sigma_1^2)^2(1+c_1\sigma_1^2)}{2\sigma_1^2(1+c_1\sigma_1^2)^2} - c_1\right)\frac{c_0}{\sqrt{2\pi}\sigma_1}\frac{\sqrt{2\pi}\sigma_1}{\sqrt{1+c_1\sigma_1^2}}$$

$$= \exp\left(\frac{\sigma_1^2(r_i - c_1)^2}{2(1+c_1\sigma_1^2)} - c_1\right)\frac{c_0}{\sqrt{1+c_1\sigma_1^2}}.$$

If we turn c_0 and c_1 back and expand $\sum_{k=1}^{r_i+1}\eta(W_{i1k})$ to $\sum_{k=1}^{r_i}\eta(W_{i1k}) + \eta(C^w_{i1(r_i+1)})$, this is the closed form of approximate denominator.

$$\exp\left(\frac{\sigma_1^2(r_i - c_1)^2}{2(1+c_1\sigma_1^2)} - c_1\right)\frac{c_0}{\sqrt{1+c_1\sigma_1^2}}$$

$$= \exp\left(\frac{\sigma_1^2\left(r_i - \sum_{k=1}^{r_i+1}\eta(W_{i1k})\right)^2}{2\left(1+\sigma_1^2\sum_{k=1}^{r_i+1}\eta(W_{i1k})\right)} - \sum_{k=1}^{r_i+1}\eta(W_{i1k})\right)\left(1+\sigma_1^2\sum_{k=1}^{r_i+1}\eta(W_{i1k})\right)^{-\frac{1}{2}}\prod_{k=1}^{r_i}\lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k})$$

$$= \exp\left(\frac{\sigma_1^2\left(r_i - \sum_{k=1}^{r_i}\eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)})\right)^2}{2\left(1+\sigma_1^2\sum_{k=1}^{r_i}\eta(W_{i1k}) + \sigma_1^2\eta(C^w_{i1(r_i+1)})\right)} - \sum_{k=1}^{r_i}\eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)})\right)$$

$$\times\left(1+\sigma_1^2\sum_{k=1}^{r_i}\eta(W_{i1k}) + \sigma_1^2\eta(C^w_{i1(r_i+1)})\right)^{-\frac{1}{2}}\prod_{k=1}^{r_i}\lambda_{01}(W_{i1k})\exp(\boldsymbol{\beta}_1^T\boldsymbol{Z}_{i1k}).$$

For numerator in 3.5, it can be expressed as follows:

$$\int_{u_{i1}} \Big[S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) - S_{i1(r_i+1)}(C^w_{i1(r_i+1)} + \tau) \Big] \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k}) \Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$= \int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k}) \Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$- \int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)} + \tau) \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k}) \Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}.$$

The first integration is equal to denominator in 3.5. Also, we can easily obtain the second one by switching $C^w_{i1(r_i+1)}$ with $C^w_{i1(r_i+1)} + \tau$ from the denominator. Hence, we approximate the numerator through the similar process above as follows:

$$\int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)}) \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k}) \Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$- \int_{u_{i1}} S_{i1(r_i+1)}(C^w_{i1(r_i+1)} + \tau) \prod_{k=1}^{r_i} \Big(- S'_{i1k}(W_{i1k}) \Big)^{\delta_{i1k}} f_{u_{i1}}(u_{i1}|\sigma_1) du_{i1}$$

$$\approx \exp\left(\frac{\sigma_1^2 \Big(r_i - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)}) \Big)^2}{2\Big(1 + \sigma_1^2 \sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)}) \Big)} - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)}) \right)$$

$$\times \Big(1 + \sigma_1^2 \textstyle\sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)}) \Big)^{-\frac{1}{2}} \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k})$$

$$- \exp\left(\frac{\sigma_1^2 \Big(r_i - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)} + \tau) \Big)^2}{2\Big(1 + \sigma_1^2 \sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)} + \tau) \Big)} - \sum_{k=1}^{r_i} \eta(W_{i1k}) - \eta(C^w_{i1(r_i+1)} + \tau) \right)$$

$$\times \Big(1 + \sigma_1^2 \textstyle\sum_{k=1}^{r_i} \eta(W_{i1k}) + \sigma_1^2 \eta(C^w_{i1(r_i+1)} + \tau) \Big)^{-\frac{1}{2}} \prod_{k=1}^{r_i} \lambda_{01}(W_{i1k}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{Z}_{i1k}).$$

BIBLIOGRAPHY

[1] Ahmed Allam, Mate Nagy, George Thoma, and Michael Krauthammer. Neural networks versus logistic regression for 30 days all-cause readmission prediction. *Scientific reports*, 9(1):1–11, 2019.

[2] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[3] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[4] NE Breslow. Discussion of the paper by cox, dr, journal of the royal statistical society. *Series B*, 34:216–217, 1972.

[5] Jenna Collins, Ibrahim M Abbass, Raymond Harvey, Brandon Suehs, Claudia Uribe, Jonathan Bouchard, Todd Prewitt, Tony DeLuzio, and Elsie Allen. Predictors of all-cause 30 day readmission among medicare patients with type 2 diabetes. *Current Medical Research and Opinion*, 33(8):1517–1523, 2017.

[6] Richard J Cook, Edmund TM Ng, Jayanti Mukherjee, and David Vaughan. Two-state mixed renewal processes for chronic disease. *Statistics in Medicine*, 18(2):175–188, 1999.

[7] Richard John Cook, Jerald F Lawless, et al. *The statistical analysis of recurrent events*. Springer, 2007.

[8] Thomas G Dietterich et al. Ensemble learning. *The handbook of brain theory and neural networks*, 2(1):110–125, 2002.

[9] Luc Duchateau and Paul Janssen. *The frailty model*. Springer, 2008.

[10] Jarrod D Frizzell, Li Liang, Phillip J Schulte, Clyde W Yancy, Paul A Heidenreich, Adrian F Hernandez, Deepak L Bhatt, Gregg C Fonarow, and Warren K Laskey. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA cardiology*, 2(2):204–209, 2017.

[11] Vivien Goepp, Olivier Bouaziz, and Grégory Nuel. Spline regression with automatic knot selection. *arXiv preprint arXiv:1808.01770*, 2018.

[12] Sara Bersche Golas, Takuma Shibahara, Stephen Agboola, Hiroko Otaki, Jumpei Sato, Tatsuya Nakae, Toru Hisamitsu, Go Kojima, Jennifer Felsted, Sujay Kakarmath, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC medical informatics and decision making*, 18(1):1–17, 2018.

[13] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[14] Xuelin Huang and Lei Liu. A joint frailty model for survival and gap times between recurrent events. *Biometrics*, 63(2):389–397, 2007.

[15] Yinan Huang, Ashna Talwar, Satabdi Chatterjee, and Rajender R Aparasu. Application of machine learning in predicting hospital readmissions: a scoping review of the literature. *BMC medical research methodology*, 21(1):1–14, 2021.

[16] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*. John Wiley & Sons, 2011.

[17] Kaitlin Kirasich, Trace Smith, and Bivin Sadler. Random forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Science Review*, 1(3):9, 2018.

[18] JF Lawless and DYT Fong. State duration models in clinical and observational studies. *Statistics in Medicine*, 18(17-18):2365–2376, 1999.

[19] Qing Li. *A two-stage pseudo likelihood approach to estimation and inference for alternating recurrent events data*. PhD thesis, University of Iowa, 2018.

[20] Yimei Li, E Paul Wileyto, and Daniel F Heitjan. Prediction of individual long-term outcomes in smoking cessation trials using frailty models. *Biometrics*, 67(4):1321–1329, 2011.

[21] DY Lin. On the breslow estimator. *Lifetime data analysis*, 13(4):471–480, 2007.

[22] Yu-Wei Lin, Yuqian Zhou, Faraz Faghri, Michael J Shaw, and Roy H Campbell. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PloS one*, 14(7):e0218942, 2019.

[23] Lei Liu, Robert A Wolfe, and Xuelin Huang. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3):747–756, 2004.

[24] Audrey Mauguen, Bernard Rachet, Simone Mathoulin-Pélissier, Gaetan MacGrogan, Alexandre Laurent, and Virginie Rondeau. Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Statistics in medicine*, 32(30):5366–5380, 2013.

[25] Yassin Mazroui, Simone Mathoulin-Pélissier, Gaetan MacGrogan, Véronique Brouste, and Virginie Rondeau. Multivariate frailty models for two types of recurrent events with a dependent terminal event: application to breast cancer data. *Biometrical Journal*, 55(6):866–884, 2013.

[26] Xu Min, Bin Yu, and Fei Wang. Predictive modeling of the hospital readmission risk from patients' claims data using machine learning: a case study on copd. *Scientific reports*, 9(1):1–10, 2019.

[27] Cécile Proust-Lima and Jeremy MG Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10(3):535–549, 2009.

[28] Bhargava K Reddy and Dursun Delen. Predicting hospital readmission for lupus patients: An rnn-lstm-based deep-learning methodology. *Computers in biology and medicine*, 101:199–209, 2018.

[29] Samuli Ripatti and Juni Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022, 2000.

[30] Dimitris Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829, 2011.

[31] Virginie Rondeau, Simone Mathoulin-Pelissier, Hélène Jacqmin-Gadda, Véronique Brouste, and Pierre Soubeyran. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*, 8(4):708–721, 2007.

[32] Daniel J Rubin. Correction to: hospital readmission of patients with diabetes. *Current diabetes reports*, 18(4):1–9, 2018.

[33] Daniel J Rubin and Arnav A Shah. Predicting and preventing acute care re-utilization by patients with diabetes. *Current Diabetes Reports*, 21(9):1–13, 2021.

[34] Pouya Saeedi, Inga Petersohn, Paraskevi Salpea, Belma Malanda, Suvi Karuranga, Nigel Unwin, Stephen Colagiuri, Leonor Guariguata, Ayesha A Motala, Katherine Ogurtsova, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas. *Diabetes research and clinical practice*, 157:107843, 2019.

[35] Sheojung Shin, Peter C Austin, Heather J Ross, Husam Abdel-Qadir, Cassandra Freitas, George Tomlinson, Davide Chicco, Meera Mahendiran, Patrick R Lawler, Filio Billia, et al. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. *ESC heart failure*, 8(1):106–115, 2021.

[36] Jade Gek Sang Soh, Wai Pong Wong, Amartya Mukhopadhyay, Swee Chye Quek, and Bee Choo Tai. Predictors of 30-day unplanned hospital readmission among adult patients with diabetes mellitus: a systematic review with meta-analysis. *BMJ Open Diabetes Research and Care*, 8(1):e001227, 2020.

[37] Maneesh Sud, Bing Yu, Harindra C Wijeysundera, Peter C Austin, Dennis T Ko, Juarez Braga, Peter Cram, John A Spertus, Michael Domanski, and Douglas S Lee. Associations between short or long length of stay and 30-day readmission and mortality in hospitalized patients with heart failure. *JACC: Heart Failure*, 5(8):578–588, 2017.

[38] Lili Wang, Kevin He, and Douglas E Schaubel. Penalized survival models for the analysis of alternating recurrent event data. *Biometrics*, 76(2):448–459, 2020.

[39] Sijin Wen, Xuelin Huang, Ralph F Frankowski, Janice N Cormier, and Peter Pisters. A bayesian multivariate joint frailty model for disease recurrences and survival. *Statistics in medicine*, 35(26):4794–4812, 2016.

[40] Xiaonan Xue and Ron Brookmeyer. Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis*, 2(3):277–289, 1996.