Risk Prediction of Opioid Use Disorder (OUD) using Electronic Health Records

Wenqiu Cao
University of Rhode Island, wenqiu_cao@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation
https://digitalcommons.uri.edu/oa_diss/1443

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.
RISK PREDICTION FOR OPIOID USE DISORDER (OUD) USING ELECTRONIC HEALTH RECORDS

BY

WENQIU CAO

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2022
DOCTOR OF PHILOSOPHY DISSERTATION
OF
WENQIU CAO

APPROVED:
Dissertation Committee:

Major Professor    Theodore A. Walls
                   Lynda A.R. Stein
                   Yichi Zhang
                   Jing Wu
                   Xuerong Wen
                   Brenton DeBoef

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2022
ABSTRACT

The opioid epidemic has emerged as a public health crisis, attracting increasing nationwide attention. Electronic health records (EHRs) provide rich resources to investigate and predict the risk of opioid use disorder (OUD) in real-world settings due to its diversity of data types and wide range of information. In this dissertation, I conducted three studies to investigate the association between OUD and different factors in EHR, including demographics, comorbidity, laboratory test results, medications, and opioid prescription, and develop predictive models to predict OUD risk using features from EHR data.

In Manuscript 1, we used penalized logistic regression models to predict OUD in the emergency department in order to handle the large number of predictors and imbalanced classes for OUD in EHR data. We presented the prediction performance of Lasso logistic regression, Firth logistic regression, Firth logistic regression with intercept-correction (FLIC), and Firth logistic regression with added covariate (FLAC) and show how physical and mental comorbidity contributed to the risk of opioid misuse in the emergency department.

In Manuscript 2, a shared parameter joint model for longitudinal and time-to-event data was built to investigate the association between longitudinal opioid prescription dosages and time to OUD onset after patients’ first opioid prescription from emergency department. Results from the models suggested a weak positive association between longitudinal opioid prescription dosage and the OUD onset. We also tested how the shared parameter model can handle data missing at random in a simulation study shown in Appendix B.

In Manuscript 3, we proposed a conditional Gated Recurrent Unit with decay rate (GRU-D) model to predict the risk of opioid dependence and abuse using both static features, like demographics and disease history, and temporal features,
such as laboratory test results during the entire visit. The GRU-D model allows us to capture the patterns of temporal features even though the measurements in EHR are collected irregularly or missing due to practical issues. We presented and discussed the predictive performance of our proposed conditional GRU-D with a GRU-D model with only temporal features and a GRU-D model with static features added at first time step. In addition, we investigated the feature importance using Leave-One-Covariate-Out (LOCO) approach. The top 15 most important predictors was presented, covering static features, such as insurance type, race, anxiety history, and also temporal features, such as blood test results, and medication use.
ACKNOWLEDGMENTS

I would like to express my gratitude to all those who helped me while pursuing this degree. These five years have been full of challenges and changes.

My deepest appreciation goes first to my major professor, Dr. Theodore (Ted) Walls, for his constant encouragement in my study. He supported me to find the goal of my research, and trusted me as I continued to explore my research and future career path. In addition, I would love to thank my committee members, Dr. Lynda (Lyn) Stein, Dr. Xuerong Wen, Dr. Yichi Zhang, and Dr. Jing Wu, for their support during all stages of doing my dissertation. Writing a Ph.D. dissertation can be a solo battle. Thanks to their insightful feedback and advice, I feel that I am not alone in this battle.

I thank all of the University of Rhode Island faculty, who taught my classes, worked with me, or helped me with my graduate study. I also would like to thank my friends, and cohorts at University of Rhode Island for helping me both in my study and my life in Rhode Island. Immense gratitude as always to Yiqiao Xin for his love, patience and support.

Above all, my sincere appreciation goes to my Mom for her support and brief in me during this long process. This accomplishment would not have been possible without her. I learned to be strong and independent from her.

One of my favorite movies in 2021 is "C’mon C’mon". There is one sentence, saying "whatever you planned on happening, that does not happen. Other stuff you never thought of happens. So just - c’mon, c’mon, c’mon ...” No matter how much many of us want to have our lives under control, things happen, such as an unexpected result in our study. Life happens. We just have to roll with the punches. I hope I can keep the memories of these 5 years in mind and face whatever will happen in my future life with calmness and courage.
TABLE OF CONTENTS

ABSTRACT ................................................................. ii
ACKNOWLEDGMENTS ....................................................... iv
TABLE OF CONTENTS ..................................................... v
LIST OF FIGURES ........................................................... viii
LIST OF TABLES ............................................................. ix

MANUSCRIPT

1 Evaluating Risk of Opioid Use Disorder in Electronic Health Records using Penalized Logistic Regression ............................. 1
   1.1 Introduction ......................................................... 1
   1.2 Methodology ....................................................... 5
      1.2.1 Research Design and Data Source .......................... 5
      1.2.2 Study Variables ............................................... 5
      1.2.3 Statistical Analysis ......................................... 6
      1.2.4 Penalized logistic regression model ....................... 6
   1.3 Results ............................................................. 9
      1.3.1 Demographic and medical characteristics .................. 9
      1.3.2 Penalized Logistic Regression of OUD ..................... 12
      1.3.3 Model Performance Comparison ............................. 14
   1.4 Discussion ....................................................... 17
   1.5 Conclusion ....................................................... 19
List of References ..................................................... 20

2 Shared Parameter Model of Longitudinal and Time-to-Event Data with Missing Time-varying Covariates: An Application to Opioid Use Disorder ............................. 24
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>24</td>
</tr>
<tr>
<td>2.2</td>
<td>Methodology</td>
<td>28</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Motivating Data</td>
<td>28</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Longitudinal Submodel</td>
<td>30</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Time-to-event Submodel</td>
<td>31</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Shared Parameter Framework for the Joint Model</td>
<td>32</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Missing Data</td>
<td>33</td>
</tr>
<tr>
<td>2.3</td>
<td>Results</td>
<td>34</td>
</tr>
<tr>
<td>2.4</td>
<td>Discussion</td>
<td>36</td>
</tr>
<tr>
<td>2.5</td>
<td>Conclusion</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>List of References</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>Predicting Risk of Opioid Dependence and Abuse using Deep Learning</td>
<td>45</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>45</td>
</tr>
<tr>
<td>3.2</td>
<td>Methodology</td>
<td>49</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Data source and preprocessing</td>
<td>49</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Gated Recurrent Unit-D Model</td>
<td>50</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Combining static and temporal features</td>
<td>52</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Model Development and Evaluation</td>
<td>53</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Feature Importance</td>
<td>54</td>
</tr>
<tr>
<td>3.3</td>
<td>Results</td>
<td>55</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Descriptive statistics and missingness</td>
<td>55</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Model Performance</td>
<td>56</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Risk factors based on feature importance</td>
<td>57</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>3.5 Conclusion</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>List of References</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td><strong>APPENDIX</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code for shared parameter model in Manuscript 2</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>A.1 R code for shared parameter joint model in JAGS</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Simulation for shared parameter model with missing time-varying covariates in Manuscript 2</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>B.1 Simulation</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>B.2 R code for simulating longitudinal and time-to-event data using simjm function</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>B.3 Simulation Results</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>List of References</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td><strong>BIBLIOGRAPHY</strong></td>
<td>78</td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>ROC curves for 4 models with AUC</td>
<td>16</td>
</tr>
<tr>
<td>1.2</td>
<td>Precision-recall curves for 4 models</td>
<td>17</td>
</tr>
<tr>
<td>2.1</td>
<td>Trajectory of opioid prescription dosage from 20 random patients</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>Trace plots of selected parameters from model with imputed data</td>
<td>37</td>
</tr>
<tr>
<td>2.3</td>
<td>Trace plots of selected parameters from full joint model</td>
<td>38</td>
</tr>
<tr>
<td>2.4</td>
<td>Trace plots of selected parameters from reduced joint model</td>
<td>39</td>
</tr>
<tr>
<td>3.1</td>
<td>Illustration of GRU model</td>
<td>52</td>
</tr>
<tr>
<td>3.2</td>
<td>Histogram for randomly selected three features</td>
<td>56</td>
</tr>
<tr>
<td>3.3</td>
<td>Model performance based on the area under the receiver operating characteristic curve (AUROC)</td>
<td>57</td>
</tr>
<tr>
<td>3.4</td>
<td>Model performance based on the area under the precision-recall curve (AUPRC)</td>
<td>58</td>
</tr>
<tr>
<td>3.5</td>
<td>Importance of the features in the best-performed model (GRU-D model with static added at first time step) based on LOCO</td>
<td>71</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

1.1 Demographic and medical characteristics ........................................ 10

1.2 Coefficients and odds ratio of predictors in Lasso logistic regression 13

1.3 Coefficients with standard errors (se) in Firth models ................. 14

1.4 Odds ratios with 95% confidence intervals (CIs) in Firth models 15

1.5 Model Performance Metrics ...................................................... 16

2.1 Parameter estimates for last observation imputed opioid prescription .................................................. 35

2.2 Parameter estimates for opioid prescription with missing values 35

2.3 Parameter estimates for opioid prescription in reduced joint model 36

2.4 Comparison of Deviance Information Criteria (DIC) .................. 36

3.1 Demographic characteristics .................................................... 69

3.2 Model performance based on mean AUROC, AUPRC, and F1-score ................................................................. 70

B.1 Parameter estimates for complete dataset .............................. 76

B.2 Parameter estimates for dataset with 20% missing values (MNAR) 76

B.3 Parameter estimates for dataset with 40% missing values (MNAR) 77
Evaluating Risk of Opioid Use Disorder in Electronic Health Records using Penalized Logistic Regression

1.1 Introduction

Over the past two decades, the opioid epidemic has emerged as a public health crisis, attracting increasing nationwide attention. In 2019, over 10 million Americans reported misuse of prescription opioids (SAMHSA, 2020), and opioid overdose related deaths have increased from less than 10,000 in 1999 to about 50,000 (Mattson et al., 2021; NIDA, 2021). What’s worse, the COVID-19 pandemic has sped up the increase in drug overdose deaths. Throughout the 12-month period ending in May 2020, opioid overdoses killed around five times more people than in the 12 months of 1999 (CDC and NCHS, 2020), which caused an overwhelming economic burden on national health care. A major contributing cause of opioid overdose and deaths in the United States is opioid use disorder (OUD).

The development of Opioid Use Disorder (OUD) results from various risk factors. Psychiatric disorders are associated with a greater risk of opioid use disorder. A national wide study using 2017 and 2018 National Survey on Drug Use and Health suggested that psychological distress was one of the significant predictors of OUD (Montiel Ishino et al., 2020). Edlund et al. (Edlund et al., 2007) also found that a history of mental health diagnosis was a moderate predictor of opioid misuse among veterans with chronic pain. Major depressive episode experience was also identified as one of the strongest predictors for adolescent opioid misuse in Han et al.’s machine learning models (Han et al., 2020). Another important risk factor for OUD is the personal history of illicit substance use (Turk et al., 2008; Webster, 2017; Han et al., 2020). Turk et al. reviewed multiple articles that address predictors for opioid abuse among chronic pain patients and concluded that
a history of non-opioid substance abuse is one of the most consistent predictors for opioid misuse (Turk et al., 2008).

Besides psychiatric disorders, previous studies on predicting opioid misuse have unraveled the importance of physical health conditions in the development of OUD (Webster, 2017; Ciesielski et al., 2016; Michna et al., 2004; Ives et al., 2006; Katz et al., 2013; Bilal et al., 2019). Katz et al. examined the relationship between opioid abuse and physical conditions using a population-based national sample and found that the increasing number of physical conditions, such as arteriosclerosis or hypertension, cardiovascular disease, arthritis, and any assessed medical condition, are significant risk factors for opioid abuse after controlling for social factors and mental disorders (Katz et al., 2013). In addition, hypertension, hyperlipidemia, and heart failure also persisted as notable predictors for OUD in machine learning models (Segal et al., 2020; Lo-Ciganic et al., 2020). Bilal et al. (Bilal et al., 2019) also showed that the more severe the disease, the higher the risk of OUD.

In addition, the wide sites of opioid receptors across human tissues and organ systems, including vital centers in the brain, distinguish opioids from other substances of abuse. It was shown that OUD may affect neuroendocrine functioning and immune response, which increases the risk of complications and infections, as well as the mortality of overdose (Brick, 2012). Bogdanowicz et al. also found that individuals with OUD who also have co-occurring Personality Disorder or Alcohol Use Disorder have an elevated risk of all-cause mortality (Bogdanowicz et al., 2015). Therefore, it is critical to better describe the patterns of comorbidity between mental/physical illness and OUD in clinical practice to assist doctors in treatment strategies and medication management, thereby optimizing treatment outcomes and decreasing mortality.
Large scale de-identified Electronic Health Records provide a great resource for investigating the risk factors related to OUD in a complex real-world healthcare environment, which includes demographic information, diagnoses, laboratory test results, vital signs, prescriptions, and procedures data. In the past decade, EHR data have been increasingly used to develop predictive models to support decision making, including predicting in-hospital mortality (Rajkomar et al., 2018), the intensive care readmission (Rojas et al., 2018), and suicide risks (Su et al., 2020). In the current study, we consider the logistic regression model for binary outcome classification considering its interpretability.

However, the enormous number of available variables also poses challenges for classification prediction. First of all, if the sample size is limited and the number of predictors is large, logistic regression carries with it a higher risk for model overfitting. This model memorizes too many details or even noises in the training cases, which thus can potentially lead to low accuracy when evaluated on unseen cases (Dreiseitl and Ohno-Machado, 2002). Furthermore, in real-world non-specialty medical settings, the prevalence of OUD may be very small. A 3-year observational study in six health systems indicates the OUD prevalence among primary care patients was only 1.0% (Lapham et al., 2020). The National Emergency Department Sample (NEDS) dataset, which is the largest all-payer ED database in the United States, consists of approximately 234 million adults who visited an ED in 2016 and 2017. Among all the ED visits, only 1.23% were related to opioids (Langabeer et al., 2021). In Rhode Island, where the data for this study were collected, only 298.3 per 100,000 population of ED visits were opioid-related (Weiss et al., 2017). This case is called imbalanced data, which means one class has many more samples than the rest of the classes. The classification models aim to minimize the percentage of incorrectly predicted classes in the training process.
They can achieve reliable results when the sample sizes of all categories are equal or close. But when a data set is imbalanced, the classification model may have very high accuracy for the majority class but extremely low accuracy for the minority class. Especially, when the data set is unbalanced and the feature dimension is large, the maximum likelihood estimation (MLE) of logistic regression can lead to biased or infinite estimates of coefficients or convergence failure due to complete separation (Rahman and Sultana, 2017; Box, 1971; Abdullah et al., 2019).

Various methods to handle high dimensional predictors and imbalanced data have emerged in recent years. The least absolute shrinkage and selection operator (Lasso) (Tibshirani, 1996) can achieve variable selection and parameter estimation through $L_1$-norm penalty. It can set the coefficients of certain predictors to zero in order to select only meaningful predictors from a large list of predictors. In a similar manner, Firth logistic regression employs the square root of the determinant of the Fisher Information Matrix as the penalty to remove the first order bias of MLE and obtain accurate and finite estimates (Firth, 1993). Moreover, two modifications of Firth logistic regression were suggested to prevent overestimation of predicted probabilities in the case of imbalanced data (Puhr et al., 2017). First, the Firth logistic regression with intercept-correction (FLIC) considers a simple post hoc adjustment of the intercept to get accurate estimation for better predictions. The second method, Firth logistic regression with added covariate (FLAC), utilizes a ghost factor as an added covariate to distinguish between original and pseudo data. On the augmented data, the MLE may be deployed to re-calibrate the average predicted probability to the ratio of events in the original data.

The purpose of this study is to illustrate the use of penalized logistic regression in detecting OUD in the emergency department. In particular, we would like to compare the prediction performance of several different penalized logistic regression
models, including Lasso, Firth, FLIC, and FLAC. Furthermore, we investigate the significant predictors of OUD and the patterns of comorbidity obtained from these models.

1.2 Methodology
1.2.1 Research Design and Data Source

The current study is a prognostic study with a retrospective cohort design. 231 adult patients’ electronic medical records were extracted from the EHR system at Roger Williams Medical Center between February and July of 2020. The dataset includes patients’ demographics, physical conditions at admission, and comorbidity. The patient’s OUD diagnosis was detected using ICD-10 codes (dsm, 2017), including F11.10 (Opioid Use Disorder, Mild), F11.11 (Opioid Use Disorder, Mild, in early or sustained remission), F11.20 (Opioid Use Disorder, Moderate), F11.21 (Opioid Use Disorder, Moderate, in early or sustained remission), F11.20 (Opioid Use Disorder, Severe), and F11.21 (Opioid Use Disorder, Severe, in early or sustained remission). Patients with one or more of these codes were considered OUD patients. The dataset consists of 14 patients with OUD diagnosis.

1.2.2 Study Variables

The demographics of subjects such as age, gender, race, ethnicity, body mass index (BMI), and type of insurance were included as predictors. In addition, we also consider some measurement of subjects’ physiologic conditions, such as their Glasgow Coma Score (GCS) at admission, time on mechanical ventilation, length of stay in the intensive care unit (ICU). The top 30 of most frequent diseases were identified based on patients’ comorbidity ICD-10 codes as listed in Table 1, and included in the penalized logistic regression models. The continuous variables were standardized before included in the model.
1.2.3 Statistical Analysis

The data analysis and modeling were performed using R programming language (R Core Team, 2018). The data were first checked and cleaned for outliers. Missing values were only found in the variable BMI, and were imputed with the mean of other observed values. The preliminary analysis included descriptive statistics such as frequency, mean and correlation to assess the distribution of the data.

1.2.4 Penalized logistic regression model

In binary logistic regression, a single outcome $y_i (i = 1, \cdots, n)$ follows a Bernoulli probability function that takes either 1 with probability $\pi_i$ or 0 with probability $1 - \pi_i$. The probability function can be presented through a linear function of the predictors $x_i = (1, x_{i1}, \cdots, x_{ip})$:

$$P(y_i = 1| x_i) = \pi_i = \frac{1}{[1 + exp(-x_i^T \beta)]},$$

where $\beta = (\beta_0, \beta_1, \cdots, \beta_p)^T$ includes the intercept $\beta_0$ and the unknown coefficients of the $p$ predictors we want to estimate. The MLE gives the estimate for $\beta$ by maximizing the log-likelihood function:

$$l(\beta) = \sum_{i=1}^{n} [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)].$$

The idea of penalized regression is to modify the log-likelihood function by adding a penalty function as a size constraint on the coefficients (Hastie et al., 2009). In this study, we used Lasso penalty function and Firth’s penalty function.

Lasso

To handle high dimensional predictors, Lasso logistic regression considers a $L_1$ norm penalty function (Tibshirani, 1996). Thus, the log-likelihood function
becomes:

\[ l_{\text{Lasso}}(\beta) = l(\beta) - \lambda \sum_{j=1}^{p} |\beta_j|, \]

where \( \lambda \) is a tuning parameter. We can use 10-fold cross validation to get the optimal \( \lambda \) that minimizes the log-likelihood function across all 10 partitions. Lasso logistic regression gives a sparse model where some coefficients are forced to be 0. This enables us to perform automatic variable selection and parameter estimation simultaneously, and also reduce the risk of overfitting.

**Firth**

Firth introduced the Jeffreys invariant prior as a penalty term to remove the first-order term in the asymptotic bias expansion of MLE (Firth, 1993). The Jeffreys invariant prior is the square root of the determinant of the Fisher information matrix \( |I(\beta)|^{1/2} \). The Firth’s penalised log-likelihood function is:

\[ l_{\text{Firth}}(\beta) = l(\beta) + \frac{1}{2} \log(|I(\beta)|). \]

When \( \beta = 0 \), \( \pi_i \) is maximized to 0.5, and thus Firth’s penalty term is maximized. Therefore, using this penalty term can shrink the coefficients towards 0. We can use backward elimination to select significant predictors.

However, because the Firth’s penalty gets to its maximum when \( \pi_i = 0.5 \), the effects of this penalty have a tendency to bring the predicted probability closer to one-half in comparison with MLE (Puhr et al., 2017). Thus, when there is a minority group in the data, Firth logistic regression is prone to overestimate predictions. Puhr et al. proposed two modifications of Firth logistic regression to avoid biased average predicted probabilities.

First, the Firth logistic regression with intercept-correction (FLIC) estimates the intercept and coefficients following the steps below:

1. Estimate predictors’ coefficients \( \hat{\beta}_{FL} \) by Firth’s penalization, excluding the
2. Obtain the MLE of the intercept \( \hat{\beta}_0 \) in the logistic regression \( P(y_i = 1) = (1 + \exp(-\beta_0 - x_i^T \hat{\beta}_{FL}))^{-1} \) by including \( \hat{\beta}_{FL} \) as an offset

3. So, the FLIC estimate \( \hat{\beta}_{FLIC} \) is then given by Firth’s estimate \( \hat{\beta}_{FL} \) with the intercept replaced by \( \hat{\beta}_0, \hat{\beta}_{FLIC} = (\hat{\beta}_0, \hat{\beta}_{FL,1}, \cdots, \hat{\beta}_{FL,p}) \).

In the second method, Firth logistic regression with added covariate (FLAC), a ghost indicator \( g \) to discriminate between original and weighted observations is included as an additional covariate in the model and the model is fitted using the augmented data (Puhr et al., 2017). The FLAC estimate can be obtained as follows:

1. Conduct Firth logistic regression and calculate the diagonal elements \( h_i \) of the hat matrix. The hat matrix is \( W^{1/2} X (X'WX)^{-1} X' W^{1/2} \), where \( X \) is the design matrix, \( W \) is the diagonal matrix of \( \pi_i (1 - \pi_1) \), and \( I(\beta) = X'WX \)

2. Define an augmented data set by combining the original observations, the original observations weighted by \( h_i/2 \), and the original observations weighted by \( h_i/2 \) but with \( y_i \) replaced by \( 1y_i \)

3. Define a ghost indicator \( g \), where \( g = 0 \) for original observations, and \( g = 1 \) for the other two pseudo data sets

4. Then, we can get FLAC estimate \( \hat{\beta}_{FLAC} \) using MLE on the augmented data with \( g \) as an added covariate.

We implemented the Lasso logistic regression using the R package \textit{glmnet} (Friedman et al., 2010). We performed 10-fold cross validation to find optimal \( \lambda \) that can minimize the error function. The Firth logistic regression, FLIC, and FLAC were implemented using the functions from R package \textit{logistf}
(Heinze et al., 2022). Backward elimination was conducted to achieve variable selection in the Firth models. All variables left in the model after this procedure are considered selected. The data were split into 70% training set and 30% testing set. The performance of these models in testing set was reported using measures such as sensitivity, specificity, accuracy, precision, F1-score, Akaike Information Criterion (AIC), area under the receiver-operating characteristic (ROC), and Precision recall (PR) curves.

1.3 Results
1.3.1 Demographic and medical characteristics

Table 1.1 shows the demographic and medical characteristics of the sample. Overall, the sample consists of 231 (45.0% female) patients aged 18 to 96 years at admission ($M = 63.38, SD = 16.75$). Of the sample, 65.8% ($n = 152$) were White, and 77.5% ($n = 179$) were not Hispanic or Latino. 134 patients were covered by private insurance, 81 were covered by Medicare or Medicaid, and others were uninsured or unknown. The mean Glasgow Coma Score (GCS) at admission of the sample was 7.23 ($SD = 6.59$). The length of stay in ICU in the sample ranged from 1 hour to 2116 hours, with an average of 102.35 hours and the average time on mechanical ventilation of the sample was 48.35 hours. Most of the patients (70.6%) transferred out of the ICU, 13.4% discharged from the ICU and others expired inside or outside of the ICU.

The prevalence of any OUD diagnosis in the whole sample was 6.1%. All the patients had at least one of the 30 physical or mental conditions included in Table 1.1. Of all the patients, 11.7% had 1-10 conditions, 37.2% had 11-20 conditions, and 51.1% had more than 20 conditions. Gastrointestinal disease (64.1%), hypertensive disease (50.2%), hyperlipidemia (48.1%), Long term use of drug therapy (47.2%) and chronic pain (42.9%) were more common than others. For each of the 30
physical or mental conditions, any OUD was more prevalent among those with some condition than those without it. These conditions include acidosis, other substance use disorders (cannabis, stimulants, sedatives, hypnotics, and multiple drug use disorder), and nicotine dependence (ND) or personal history of ND.

Table 1.1: Demographic and medical characteristics

<table>
<thead>
<tr>
<th></th>
<th>Opioid Use Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Age at Admission</td>
<td>64.346</td>
</tr>
<tr>
<td>BMI(kg/m^2)</td>
<td>29.438</td>
</tr>
<tr>
<td>Glasgow Coma Score (GCS)</td>
<td>6.963</td>
</tr>
<tr>
<td>at Admission</td>
<td></td>
</tr>
<tr>
<td>Time on Mechanical Ventilation (hours)</td>
<td>41.968</td>
</tr>
<tr>
<td>Length of Stay in ICU (hours)</td>
<td>94.650</td>
</tr>
<tr>
<td>Numbers of Comorbidity</td>
<td>21.700</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>119</td>
</tr>
<tr>
<td>Female</td>
<td>98</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>144</td>
</tr>
<tr>
<td>Non-white/unavailable</td>
<td>73</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Not Hispanic/Latino</td>
<td>170</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>47</td>
</tr>
<tr>
<td>Hypertensive disease</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>106</td>
</tr>
<tr>
<td>Yes</td>
<td>111</td>
</tr>
<tr>
<td>Acute respiratory failure</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>144</td>
</tr>
<tr>
<td>Yes</td>
<td>73</td>
</tr>
<tr>
<td>Acidosis</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>149</td>
</tr>
<tr>
<td>Yes</td>
<td>68</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>188</td>
</tr>
<tr>
<td>Yes</td>
<td>29</td>
</tr>
<tr>
<td>Hypokalemia</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>149</td>
</tr>
<tr>
<td>Yes</td>
<td>68</td>
</tr>
<tr>
<td>Acute kidney failure</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>149</td>
</tr>
<tr>
<td>Yes</td>
<td>68</td>
</tr>
<tr>
<td>Hypo-osmolality and hyponatremia</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>155</td>
</tr>
<tr>
<td>Yes</td>
<td>62</td>
</tr>
<tr>
<td>Hyperkalemia</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>184</td>
</tr>
<tr>
<td>Yes</td>
<td>33</td>
</tr>
<tr>
<td>Type 2 diabetes mellitus</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>187</td>
</tr>
</tbody>
</table>
with renal complications
with other multiple complications | Yes | 30 | 2
Myocardial infarction type 2 | No | 161 | 11
Yes | 56 | 3
Spesis | No | 152 | 9
Yes | 65 | 5
Hypomagnesemia | No | 170 | 10
Yes | 47 | 4
Cardiovascular and ischaemic disease | No | 163 | 12
Yes | 54 | 2
Gastrointestinal disease | No | 74 | 9
Yes | 143 | 5
Hyperlipidemia | No | 107 | 13
Yes | 110 | 1
Cancer | No | 176 | 14
Yes | 41 | 0
Chronic pain | No | 123 | 9
Yes | 94 | 5
Disorders of phosphorus metabolism and phosphatases | No | 171 | 10
Yes | 46 | 4
Metabolic encephalopathy | No | 175 | 10
Yes | 42 | 4
Diseases of musculoskeletal system and connective tissues | No | 167 | 9
Yes | 50 | 5
Pneumonia | No | 159 | 10
Yes | 58 | 4
COPD | No | 170 | 9
Yes | 47 | 5
Depression | No | 187 | 11
Yes | 30 | 3
Anxiety Disorder | No | 183 | 12
Yes | 34 | 2
Other Psychosis or mental disorder | No | 191 | 11
Yes | 26 | 3
Alcohol use disorders | No | 191 | 12
Yes | 26 | 2
Other substance use disorders | No | 184 | 3
Yes | 33 | 11
Nicotine Dependence (ND) or personal history of ND | No | 144 | 3
Yes | 73 | 11
Acute posthemorrhagic anemia and anemia | No | 151 | 10
Yes | 66 | 4
Dehydration | No | 182 | 11
Yes | 35 | 3
1.3.2 Penalized Logistic Regression of OUD

Lasso Logistic Regression

We first conducted Lasso logistic regression to identify predictor significantly correlated with OUD. Through 10-fold cross validation, we selected the value that minimized the error term as optimal tuning parameter, $\lambda = 0.014$. The results of the final Lasso regression model in Table 1.2 show that OUD was significantly associated with 13 predictors out of 41 candidate predictors. Patients, who had higher GCS at admission, was non-white, expired outside of the ICU or had acidosis, acute kidney failure, depression, other substance use disorder, nicotine dependence (ND) or personal history of ND, or dehydration, were more likely to have any OUD. Other predictors were negatively associated with any OUD, including age, gastrointestinal disease, hyperlipidemia, and alcohol use disorder. Among all 13 significant predictors, other substance use disorder (cannabis, stimulants, sedatives, hypnotics, and multiple drug use disorder) was the most significant predictor, with $\beta = 1.836, OR = 6.269$. This suggests patients with other substance use disorder were 5.269 times more likely to have any OUD than those without.

Firth Logistic Regression

Next, we fitted a Firth logistic regression model with 41 predictors. The backward elimination algorithms was applied for variable selection. There were 9 variables selected as significant predictors of OUD in the final Firth logistic
Next, the FLIC and FLAC were conducted to with these 9 variables to get unbiased or less biased estimates. Table 1.3 shows the coefficient estimates with standard errors from three Firth models. Being male, not having hyperlipidemia, anxiety disorder, or alcohol use disorder, and not being covered by Medicare/Medicaid were significantly related to having any OUD. A patient with younger age or shorter time on mechanical ventilation is more likely to have any OUD. In addition, staying in the ICU for a longer time period, or having nicotine dependency (ND) or personal history of ND were significantly related to higher risk of any OUD diagnosis.

The estimates of coefficients were almost same for Firth logistic regression and FLIC, while FLAC’s estimates were slightly different. Among all three models, FLIC provided coefficient estimates with the smallest standard errors for 9 variables.

The odds ratios with 95% confidence intervals were reported in Table 1.4. The odds ratios of two positive predictors, length of stay in the ICU (\(OR_{Firth,FLIC} = 404.171, OR_{FLAC} = 440.647\)), and ND or personal history of ND (\(OR_{Firth,FLIC} = \))
1386.617, \( OR_{FLAC} = 1462.239 \)), were large, suggesting a strong association with OUD. The wide 95\% confidence intervals and large odds ratios are due to high standard deviation in length of stay in the ICU \( (SD = 219.620) \), or imbalanced OUD classes with only 6\% positive cases. The imbalanced classes of OUD also made the odds ratios for negative predictors close to 0 as shown in Table 1.4.

Table 1.3. Coefficients with standard errors (se) in Firth models

<table>
<thead>
<tr>
<th></th>
<th>Firth</th>
<th>FLIC</th>
<th>FLAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-4.473</td>
<td>-4.473</td>
<td>-4.551</td>
</tr>
<tr>
<td></td>
<td>(1.462)</td>
<td>(1.051)</td>
<td>(1.514)</td>
</tr>
<tr>
<td>Gender</td>
<td>-3.968</td>
<td>-3.968</td>
<td>-4.022</td>
</tr>
<tr>
<td></td>
<td>(1.715)</td>
<td>(1.331)</td>
<td>(1.769)</td>
</tr>
<tr>
<td>Time on Mechanical</td>
<td>-4.540</td>
<td>-4.540</td>
<td>-4.663</td>
</tr>
<tr>
<td>Ventilation</td>
<td>(1.549)</td>
<td>(1.306)</td>
<td>(1.611)</td>
</tr>
<tr>
<td>Length of Stay in ICU</td>
<td>6.002</td>
<td>6.002</td>
<td>6.088</td>
</tr>
<tr>
<td></td>
<td>(1.960)</td>
<td>(1.531)</td>
<td>(2.036)</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>-8.781</td>
<td>-8.781</td>
<td>-8.995</td>
</tr>
<tr>
<td></td>
<td>(2.775)</td>
<td>(2.301)</td>
<td>(2.873)</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>-3.961</td>
<td>-3.961</td>
<td>-4.041</td>
</tr>
<tr>
<td></td>
<td>(1.651)</td>
<td>(1.073)</td>
<td>(1.685)</td>
</tr>
<tr>
<td>Alcohol use disorder</td>
<td>-7.498</td>
<td>-7.498</td>
<td>-7.681</td>
</tr>
<tr>
<td></td>
<td>(2.826)</td>
<td>(1.943)</td>
<td>(2.911)</td>
</tr>
<tr>
<td>ND or personal history of ND</td>
<td>7.235</td>
<td>7.235</td>
<td>7.288</td>
</tr>
<tr>
<td></td>
<td>(2.269)</td>
<td>(1.342)</td>
<td>(2.323)</td>
</tr>
<tr>
<td>Medicare or Medicaid</td>
<td>-3.894</td>
<td>-3.894</td>
<td>-4.068</td>
</tr>
<tr>
<td></td>
<td>(1.913)</td>
<td>(1.425)</td>
<td>(1.984)</td>
</tr>
</tbody>
</table>

1.3.3 Model Performance Comparison

We then compared the classification performance of 4 models using testing set. In Table 1.5, Lasso logistic regression with 13 predictor had the highest accuracy of 97.1\% and the lowest AIC of -19.903. However, in the case of imbalanced data, accuracy is not the most appropriate measure for prediction performance. Because the effect of the minority but more meaningful samples is reduced in comparison to those of the majority group (Branco et al., 2016). In this study, only 6\% of the sample belong to the OUD group, so predicting the majority class, no OUD,
| Table 1.4. Odds ratios with 95% confidence intervals (CIs) in Firth models |
|--------------------------|----------------|----------------|
|                        | Firth       | FLIC          | FLAC          |
| **Age**                 | 0.011       | 0.011         | 0.011         |
|                         | (0.000, 0.172) | (0.000, 0.172) | (0.000, 0.109) |
| **Gender**              | 0.019       | 0.019         | 0.018         |
|                         | (0.000, 0.489) | (0.000, 0.489) | (0.000, 0.300) |
| **Time on Mechanical**  | 0.011       | 0.011         | 0.009         |
| **Ventilation**         | (0.000, 0.226) | (0.000, 0.226) | (0.000, 0.133) |
| **Length of Stay in ICU**| 404.171    | 404.171       | 440.647       |
|                         | (9.156, 2589446.577) | (9.156, 2589446.577) | (17.173, 65033.350) |
| **Hyperlipidemia**      | 0.000       | 0.000         | 0.000         |
|                         | (0.000, 0.030) | (0.000, 0.030) | (0.000, 0.012) |
| **Anxiety disorder**    | 0.019       | 0.019         | 0.018         |
|                         | (0.000, 0.436) | (0.000, 0.436) | (0.000, 0.282) |
| **Alcohol use disorder**| 0.001       | 0.001         | 0.000         |
|                         | (0.000, 0.131) | (0.000, 0.131) | (0.000, 0.048) |
| **ND or personal**      | 1386.617    | 1386.617      | 1462.239      |
| **history of ND**       | (19.209, 36291964.460) | (19.209, 36291964.460) | (38.178, 472565.953) |
| **Medicare or Medicaid**| 0.020       | 0.020         | 0.017         |
|                         | (0.000, 0.967) | (0.000, 0.967) | (0.000, 0.470) |

for all patients can also achieve an accuracy of 94%. But, all the patients in the minority group will be misclassified and the potential risk of OUD cannot be detected. Hence, besides accuracy and AIC, we also reported the recall (sensitivity or true positive rate), specificity (true negative rate), precision, and F1-score, which combines recall and precision (Rijsbergen, 1979). As shown in Table 1.5, Firth logistic regression, FLIC and FLAC have higher scores for recall, specificity, precision and F1-score than Lasso logistic regression. Especially, Firth logistic regression has the highest F1-score of 0.961, and predicted all the positive cases correctly. The performance of FLIC and FLAC were similar.

The receiver operating characteristics (ROC) curve and the corresponding area under the ROC curve (AUC) are also very popular in imbalance data prediction (Metz, 1978). The ROC curve visualizes of the relative trade-off between true
positive rate and false positive rate. The ROC curves with AUC for 4 models are shown in Figure 1.1. Three Firth models have similar curves, while Firth logistic regression has the highest AUC of 0.933. Overall, the Firth models outperformed Lasso logistic regression.

Precision-recall curves (PR curves) are also recommended for highly imbalanced data to evaluate model performance (Davis and Goadrich, 2006). Figure 1.2 shows the PR curves for 4 models. The curves of Firth logistic regression and FLIC overlap. At thresholds with low recall, the precision is correspondingly higher for 3 Firth models, which indicates 3 Firth models slightly outperformed the lasso model. But as recall scores increased, the precision scores became close in all 4 models, which may be caused by the extreme imbalanced classes.

<table>
<thead>
<tr>
<th>Model</th>
<th>Recall</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>Precision</th>
<th>F1-score</th>
<th>AUC</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso</td>
<td>0.000</td>
<td>0.000</td>
<td>0.971</td>
<td>0.000</td>
<td>0.000</td>
<td>0.896</td>
<td>-19.903</td>
</tr>
<tr>
<td>Firth</td>
<td>0.925</td>
<td>1.000</td>
<td>0.928</td>
<td>1.000</td>
<td>0.961</td>
<td>0.933</td>
<td>68.311</td>
</tr>
<tr>
<td>FLIC</td>
<td>0.925</td>
<td>0.500</td>
<td>0.913</td>
<td>0.500</td>
<td>0.649</td>
<td>0.933</td>
<td>67.785</td>
</tr>
<tr>
<td>FLAC</td>
<td>0.925</td>
<td>0.500</td>
<td>0.913</td>
<td>0.500</td>
<td>0.649</td>
<td>0.925</td>
<td>75.712</td>
</tr>
</tbody>
</table>

Figure 1.1. ROC curves for 4 models with AUC
1.4 Discussion

The aim of the current study was to identify significant risk factors associated with OUD and establish a prediction model for OUD diagnosis among ICU patients using Electronic Health Records. This study focused on the challenges introduced by imbalanced data with high dimensional features and utilized two types of penalized logistic regression, including Lasso and Firth logistic regression, and two modifications of Firth model as alternatives to the classical logistic regression in order to avoid overfitting and biased estimate and improve the predictive power for minority group.

This study found 19 significant risk factors associated with OUD from 4 models. Predictors shared across all 4 models include age at admission, hyperlipidemia, alcohol use disorder (AUD), and nicotine dependence (ND) or personal history of ND. Some classification studies have reported that a younger age is a significant predictor for opioid misuse (Webster, 2017; Turk et al., 2008; Ives et al., 2006; Edlund et al., 2007). Hyperlipidemia, and a history of non-opioid substance abuse are also supported as consistent predictors for OUD in two recent machine learning models (Segal et al., 2020; Lo-Ciganic et al., 2020). The strong association
between heavy smoking/nicotine dependency and opioid misuse was also found in a large representative sample from the National Survey on Drug Use and Health (Zale et al., 2014). Our study reinforced the important role of age, hyperlipidemia, AUD, and ND or a personal history of ND in the prediction of OUD.

The comparison of coefficient estimates suggests FLIC can most accurate estimates with least bias. A previous simulation study also reported FLAC may introduce some bias in the effect estimation in order to achieve a better prediction performance (Puhr et al., 2017). However, in this study, Firth logistic regression outperformed FLIC and FLAC with the highest F1-score, which may be related to the small prevalence of OUD patients in a limited sample size. Puhr et al. suggested increasing sample size and expected event rate can reduce the difference between Firth logistic regression and FLIC/FLAC (Puhr et al., 2017). Furthermore, comparing 3 Firth models with Lasso logistic regression, we found Firth models, especially the ordinary Firth logistic regression, had more power in detecting minority but meaningful cases than Lasso logistic regression.

The current study has some limitations. First, the sample was extracted from one location at one single time point, which might not be able to represent the population and did not consider the longitudinal effect of some factors, such as history of opioid medication use. Second, our sample has a limited size and the number of OUD patients was very small. Even though we applied penalized methods to handle these issues, the performance of classification models heavily depend on the amount of data used for training. So the PR curves for 4 models in our study showed poor performance. Finally, we did not utilize any resampling method to get a more balanced data for OUD prediction because we believe the original data can better characterize the pattern of OUD patients in the realistic medical settings. However, this choice also sacrificed partial accuracy of the prediction.
Given the limitations in our sample, future study should focus on increasing the sample size and the representativeness of the sample by integrating more patients’ records from multiple locations, as well as incorporating more potential risk factors, such as longitudinal records of opioid medication use, and the severity of certain physical/mental condition. To improve the prediction performance in the future, we can utilize some resampling methods to oversample the cases in the minority class, such as synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002).

Strengths of the current study involves both practical and methodological aspects. We extracted 30 physical/mental conditions based on ICD-10 code and investigated their associations with OUD in a real-world medical setting. We identified several strong predictors for clinicians to consider when assessing the risk of OUD among ICU patients. In addition, we utilized several penalized logistic regression to select significant predictors from a large amount of features and adopted Firth logistic regression and its modifications to adjust the estimate and prediction bias in the case of imbalanced classification. Our results illustrate with Firth penalization, logistic regression can classify minority cases more correctly.

1.5 Conclusion

As opioid crisis continues to grow across the U.S., efforts to provide timely intervention and better healthcare for patients at high risk of opioid misuse have become more urgent. This study sought to shed insight on predictors of OUD in a real-world sample of ICU patients, and observed that a variety of demographic, mental, behavioral and physical predictors contributed to the risk of opioid misuse. Our findings obtained in the comparison of model estimation and prediction suggest that Firth penalized logistic regression and its modifications, FLIC and FLAC, are favorable to predict rare cases in imbalanced data.
List of References

(2017). Dsm-5 diagnose and new icd-10-cm codes.


CDC and NCHS (2020). Wide-ranging online data for epidemiologic research (wonder).


SAMHSA (2020). Key substance use and mental health indicators in the United States: Results from the 2019 national survey on drug use and health (hhs publication no. pep20-07-01-001, nsdhu series h-55). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.


MANUSCRIPT 2

Shared Parameter Model of Longitudinal and Time-to-Event Data with Missing Time-varying Covariates: An Application to Opioid Use Disorder

2.1 Introduction

Electronic Health Record (EHR) data include a variety of variables that may be critical for the development of opioid use disorder (OUD), such as personal demographic information, comorbidities and disease history, medical procedures, and medication prescriptions. If the patient repeatedly attended the same facility or several facilities using the same EHR system, his/her visit records may be considered as longitudinal data. In the meanwhile, time to event data, such as the time to the onset of a particular illness, can sometimes be studied together with longitudinal and demographic information. Consequently, we may use EHR data to investigate the associations between the longitudinal history of a variable and its influence on the risk of an event. To achieve the goal of modeling both longitudinal and time-to-event data simultaneously, a class of joint model has been developed (De Gruttola and Tu, 1994; Tsiatis et al., 1995; Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997). Joint models for longitudinal and time-to-event data are a powerful method that cannot only bring these two data types together in a single model, but also take into consideration the dependence and relationship between longitudinal and time-to-event data. Previous studies show joint models have the potential to lessen the burden of bias and improve the efficiency of the estimations (Ibrahim et al., 2010; Zhang et al., 2014). Therefore, they have gained increasing attention in the statistics field and been applied in a wide range of studies in epidemiological and biomedical fields.

However, EHR data, unlike data from clinical trials used in previous joint
model studies, have some commonly presented practical issues, such as missing observations (Ward et al., 2015). This may pose challenges to accurate estimation for the longitudinal data (Ghassemi et al., 2020; Rajkomar et al., 2018). There are two categories of missing data in longitudinal studies, intermittent missing and dropout. Intermittent missing refers to an unobserved value followed by an observed value, while a missing value because of dropout has no follow-up observations (Gad and Darwish, 2013). Both are observed in EHR data. To handle missing values in longitudinal data, the underlying mechanism of missingness needs to be understood first. Missing mechanisms include missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). A missing variable is considered MCAR when the probability of missingness is independent of both observed and unobserved data, and MAR if the missingness only depends on fully observed variables. When observations are MCAR or MAR, there are three approaches we can use complete case analysis or multiple imputation to handle missing values and achieve unbiased estimates (Little and Rubin, 2019; Groenwold et al., 2012). Complete-case analysis only uses data from those subjects without any missing observations, and multiple imputation replaces the missing values with a plausible values from multiple completed datasets. These methods have been used widely in clinical trials or observational studies. But these methods are valid under the assumption of MCAR or MAR. Especially, when the degree of missing becomes large, removing incomplete cases will not only produce biased estimates but also yield a large loss of information. When the missing mechanism is neither MCAR nor MAR, the data are considered as MNAR, where the missing probability of a variable depends upon its unobserved value. Under MNAR, unbiased estimation needs to account for the probability of missing, the distribution of the values of the missing variable and the rela-
tionship between the incomplete variable and response variable. To model all three elements simultaneously, selection models (Heckman, 1976), pattern mixture models (Glynn et al., 1986) and shared parameter models (Wu and Carroll, 1988; Gad and Darwish, 2013; Vonesh et al., 2006) were introduced. In a shared parameter model, a random effect is shared between longitudinal model and the missing mechanism model. Many studies have used the shared parameter model for missing values in longitudinal data and extended it to a wide range of response types, including continuous normally distributed response (Wu and Carroll, 1988), binary response (Follmann and Wu, 1995) and count data (Albert and Follmann, 2000).

With the increasing research in joint model, there is also extensive literature on missing data in joint models. Wu et al. incorporated a missing data mechanism into the joint model likelihood to explore informative dropouts in an AIDS clinical trial, and used an expectation–maximization (EM) method inside the likelihood framework for the estimation of the model parameters (Wu et al., 2008). Thiébaut et al. constructed a joint model for the bivariate linear mixed model and survival model to account the effect of two longitudinal biomarkers, one of which was subject to left censoring, and estimated the parameters using a direct maximum likelihood approach (Thiébaut et al., 2005). Recently, Chen at al. introduced a Bayesian approach for a longitudinal study with censored and missing time-varying covariate data within the joint-modeling framework to investigate the longitudinal AIDS progression and failure from AIDS (Chen et al., 2014). In this study, we aim to model the longitudinal and time to event data jointly within a shared parameter framework. The shared parameter joint model is one of the most common joint modelling approach in the literature. A shared parameter joint model connects two related submodels, longitudinal and time-to-event submodels, through a shared individual-level random effects. Due to this feature and the
conditional independence assumptions, shared parameter models can handle both intermittent missing and dropout easily without considering the missing longitudinal observations (Rizopoulos, 2012).

The motivation for this study lies in our interest in investigating available risk factors that may be associated with the onset of opioid use disorder (OUD) among emergency department (ED) patients. Data used in this study were collected from 2001 to 2012. During this period, according to a report of opioid prescription pattern in 2009, the number of prescriptions for opioids from emergency department (EDs) was the third highest of all opioid prescription sources among the patients younger than 29 years old, around 12% of the total prescriptions, and the fourth highest in the 30-39-year-old age group (Volkow et al., 2011). Despite the fact that these findings do not examine the total doses of opioid, it can also be emphasized that the ED patients are at a high risk for opioid dependence and abuse (Hansen, 2005). Factors related to OUD, such as medical history of opioid prescribed in ED, are of particular interest. Previous observational studies suggest that opioid-naive patients prescribed first opioids in ED were at elevated risk for long-term recurrent opioid use (Hoppe et al., 2015; Barnett et al., 2017; Butler et al., 2016). In addition, Butler et al. found the time from initial exposure to onset of nonmedical use to get high ranged from 2 to 36 months, and after a median of 12 months, the patient may start using regularly to get high (Butler et al., 2016). Although the effect of opioid prescription from EDs on the risk of OUD onset remains unclear, these previous studies showed evidence suggesting a further investigation. Hence, our objective of this study is to evaluate how opioid prescription from EDs over time affect the onset of opioid use disorder. However, a challenge of using EHR data for this analysis is there is a significant portion of missing data, because it is unrealistic for everyone to come to the ED
regularly like in clinical trials, and the length of time from the first visit to the last visit varies from person to person in the system. Specifically, given the high risk of overdose associated with opioid, we suspect that the data in longitudinal opioid prescription is not MCAR. For pain management in ED, there are multiple analgesic drugs that can be as effective as opioids. Before providers prescribed opioid to a patients, they need to check patient’s medical history and conduct an assessment of the pain, physical and psychological function. So, the missingness may be depend on other variables, such as history of substance use disorder and chronic/acute pain conditions. Thus, we suspect the missing mechanism for longitudinal opioid prescription doses in our data is MAR.

As previously noted, shared parameter models may handle complex missing data and offer unbiased estimates. Hence, the main objective of this work is to employ shared parameter models to jointly analyze longitudinal opioid prescription dosages and time to OUD onset. The relationship between them is also examined within the context of the joint model. In addition, we investigate the efficacy of the shared parameter model with respect to missing longitudinal responses by comparing estimates obtained under scenarios with different proportion of missing data.

2.2 Methodology
2.2.1 Motivating Data

The current study is conducted using an open-access dataset, Medical Information Mart for Intensive Care III (Mimic-III) (Johnson et al., 2016) provided by the Massachusetts Institute of Technology. This dataset is composed of de-identified, comprehensive clinical data of patients admitted to the critical care units of the Beth Israel Deaconess Medical Center in Massachusetts between 2001 and 2012. It includes 58,976 admissions from 48,520 distinct adult patients, with 1190 adult
patients having OUD related diagnosis. The dataset has been used for different research purposes such as discharge or mortality prediction (McWilliams et al., 2019; Pirracchio et al., 2015). It should be noted that all dates in this dataset have been shifted randomly to protect patient confidentiality, which means dates are internally consistent for the same patient, but cannot represent the actual time.

In this study, we first identified patients with opioid prescription using National Drug Code (NDC) (CDC, 2021). In this data, we found two types of opioids, fentanyl citrate and morphine sulfate. Then, we selected patients with at least 5 visits. The time of patients’ first opioid prescription is used as the baseline reference time. For each patient, we calculated the interval between each of his/her subsequent visits and the reference time on a monthly basis. The event of interest is the onset of first OUD diagnosis, which is coded based on OUD diagnosis ICD-9 codes, include 304.00–304.03 (Opioid type dependence), 304.70–304.73 (Combinations of opioids with any other), 305.50–305.53 (Nondependent opioid abuse), 965.0 (Poisoning; Opiates and Related Narcotics), 965.00 (Poisoning; Opium/alkaloids, unspecified), 965.01 (Poisoning; Heroin), 965.02 (Poisoning; Methadone), 965.09 (Poisoning; Other opiates and related narcotics), E85.00 (Accidental Poisoning; Heroin), E85.01 (Accidental Poisoning; Methadone), 97.01 (Poisoning; Opiate antagonists) and E85.02 (Accidental Poisoning; Other Opiates and Related Narcotics) (Heslin et al., 2015). We considered the time from a patient’s first opioid prescription to first diagnosis for OUD. Time-independent covariates include age of first opioid prescription, chronic pain condition indicator, admission type of first opioid prescription, and insurance type.

The time-varying longitudinal variable is the total dosage of opioid prescription. Since we have two types of opioid, we transformed the dosage of fentanyl to equivalent dosage of morphine. A dose of 100 mcg fentanyl is approximately equiv-
alent in analgesic activity to 10 mg of morphine (FDA, 2013). If multiple doses of the same opioid were prescribed in the same month, we merged those prescriptions and calculated a total dosage of opioid prescription. Any values larger than 1,000 mg of morphine were excluded from the analysis as extreme outliers. After pre-processing, there are 43.9% missing intermittently in the longitudinal variable. A trajectory plot for opioid dosage for a random sample of 20 patients in the study is given in Figure 2.1. We can observe obvious individual level differences so we used a mixed model to account for individual level random effect in longitudinal variable.

![Figure 2.1. Trajectory of opioid prescription dosage from 20 random patients](image)

### 2.2.2 Longitudinal Submodel

Longitudinal data consist of repeated measurements for each individual over time. Each individual in the sample is expected to have his/her own individual level pattern over time. This between-person variability is accounted for by the mixed-effects model, which estimates person-specific random effects around parameters that are fixed across persons (Laird and Ware, 1982).

Let \( y_i(t) \) denote a longitudinal variables observed at the time \( t(t = 1, \ldots, T_i^*) \) for the \( i^{th} (i = 1, \ldots, N) \) individual in a sample of \( N \) individuals, where \( T_i^* \) is the time-to-event for individual \( i \). We specify a linear mixed model for the longitudinal opioid prescription dosage as:
\[ y_i(t) = X_i(t)\beta + Z_i(t)b_i + \epsilon_i(t) \]  \hspace{1cm} (2.1)

where \( X_i(t) \) and \( Z_i(t) \) represent the design matrix containing the predictors for the fixed effects regression coefficients \( \beta \) and for the random effects regression coefficients \( b_i \), respectively. The random effects \( b_i = (b_{i0}, b_{i1})' \) are for random intercept and slope effects. The measurement error term \( \epsilon_i(t) \) is the error term for individual \( i \) at time \( t \). All the error terms are assumed mutually independent, independent of \( b_i \), and normally distributed with mean zero and variance \( \sigma^2 \).

### 2.2.3 Time-to-event Submodel

Time-to-event analysis, also known as survival analysis, refers to statistical approaches to analyze time-to-event data. An event time is the length of time that has passed until an event of interest, such as a death or a disease diagnosis, happened. In time-to-event data, when the event of interest might not occur for every individual within the study period, or an individual drops out before the end of the study, this type of missing data is treated as right censored data. There are several different models for time-to-event analysis (Kaplan and Meier, 1958). Cox proportional hazard (PH) model is one of the most popular models, where the multiplicative effect of covariates on the hazard for an event is modeled in a regression (Cox, 1972). The baseline hazard function in COX PH model has no exact pre-specified form and is estimated non-parametrically. We specify a COX PH model for the time to OUD onset as:

\[ h_i(t) = h_0(t)exp(\gamma^T w_i), \]  \hspace{1cm} (2.2)

where \( h_0(t) \) is the baseline hazard function at time \( t \) and \( w_i \) is a vector of baseline explanatory covariates with corresponding coefficients \( \gamma \).
2.2.4 Shared Parameter Framework for the Joint Model

The longitudinal and time-to-event submodels can be connected through shared random-effects (Wulfsohn and Tsiatis, 1997). Under this framework, only the random effects from the longitudinal submodel in Equation 2.1 are included in the relative risk model. The random effects (intercept and slope) represent the individual departures from the sample mean. Therefore, the association parameters used to interconnect the random effect and relative risk models reflect the change in log hazard for a one-unit change in these deviations. Based on this shared parameter model, we can see how an individual’s baseline observation of the longitudinal data (random intercept) and longitudinal trend (random slope) impact his/her probability of encountering the event of interest.

The joint model for longitudinal and time-to-event model assumes a full joint distribution for the longitudinal responses, survival process and random effects as (Armero et al., 2018):

$$p(y_i, T_i, \delta_i) = \int p(y_i|b_i)\{h(T_i|b_i)^{\delta_i}S(T_i|b_i)\}p(b_i)db_i,$$

(2.3)

where $y_i$ is the longitudinal variable, $T_i$ is the observed event time for patient i, and $\delta_i$ is the event indicator. For the survival process, the survival function for patient i is $S(T_i|b_i)$ and the relative risk model is $h(T_i|b_i)$. The random effects $b_i$ explain all interdependencies.

Here, we define a set of conditional independence assumptions for the shared parameter joint model, that is: (i)$y_i(t) \perp T^*_i|b_i$; (ii)$y_i(t) \perp y_i(t')|b_i$. Assumption (i) implies that the longitudinal outcome is independent of the time-to-event outcome, while Assumption (ii) assumes that the repeated measurements in the longitudinal outcome are independent of each other. Then, with the two submodels we defined
in previous section, we can specify the joint model as:

\[ y_i(t) = X_i(t)\beta + Z_i(t)b_i + \epsilon_i(t), \]

\[ h_i(t|h_0, \gamma, \alpha, b_i) = h_0(t)\exp(\gamma^T w_i + \alpha Z_i(t)b_i), \]

where \( \alpha \) is the association parameter, and \( Z_i(t)b_i \) is the random effect. And the survival function for individual \( i \) is defined as:

\[ S_i(t|h_0, \gamma, \alpha, b_i) = \exp\{-\int_0^t h_i(s|h_0, \gamma, \alpha, b_i)ds\}, \]

which can be approximated using Gaussian quadrature method (Rizopoulos, 2012).

We built the model in R (R Core Team, 2018) using \textit{rjags} (Lunn et al., 2009) and \textit{runjags} (Denwood, 2016). The model defined in JAGS is shown in Appendix A.1. In terms of the Gaussian quadrature method used for integration estimation, we used the package \textit{statmod} (Giner and Smyth, 2016).

2.2.5 Missing Data

Shared parameter framework allows the joint model for longitudinal and time-to-event data to handle missing values easily given the conditional independence assumptions. Let \( y_i^o \) denote the observed longitudinal responses, and \( y_i^m \) denote the unobserved missing values. Then, the log likelihood under the complete data model \( \{y_i^o, y_i^m\} \) for longitudinal data is defined as:

\[
\begin{align*}
    l(\theta) &= \sum_{i=1}^N \log \int p(T_i, \delta_i, y_i^o, y_i^m; \theta)dy_i^m \\
    &= \sum_{i=1}^N \log \int p(Y_i, \delta_i|b_i; \theta)\{\int p(y_i^o, y_i^m|b_i; \theta)dy_i^m\}p(b_i; \theta)db_i 
\end{align*}
\]

Under Assumption (i), the missing longitudinal responses \( y_i^m \) are only involved in the density function of the longitudinal submodel. In addition, under Assumption (ii), the longitudinal responses conditionally on the random effects are
independent with each other, so the integral of \( y_i^m \) in Equation 2.4 can be dropped. Then, we have the log likelihood specified as:

\[
    l(\theta) = \sum_{i=1}^{N} \log \int p(Y_i; \delta_i; b_i; \theta)p(y_i^0|b_i; \theta)p(b_i; \theta) db_i,
\]  

(2.5)

which can be obtained without considering the integration with respect to the missing at random responses.

A simulation study was conducted to evaluate the performance of the shared parameter joint model regarding the percentage of missing values in Appendix B.

2.3 Results

As described before, we included 4 time-independent covariates, including age of first opioid prescription, chronic pain condition indicator, admission type of first opioid prescription, and insurance type in the time-to-event submodel. For the longitudinal dosage of opioid prescription, we only included fixed and random effects of time. Another strategy to handle missing values was also used, imputation using the last available observation, to create a complete dataset. Shared parameter joint models were fit for the complete dataset and the original dataset separately using 3 chains of 10,000 MCMC iterations, which included 1,000 burn-in iterations. The estimates for the parameters are reported in Table 2.4 and 2.5.

From Table 2.4 and 2.5, the potential scale reduction factor (psrf) scores are close to 1, indicating that each of the simulated observations is close to the target distribution. By comparing the estimated in two models, we can find that imputing missing opioid prescription with last available observation increases the mean standard deviation (SD) of the posterior distribution and may be not be able to reflect the real distribution of the data. In addition, in the model using imputed data, none of the covariates are significant predictors for the event time. In the shared parameter joint model with missing values, the SDs are smaller than those in the imputed model. In addition, the association parameter, \( \alpha \), is 0.041 with
Table 2.1. Parameter estimates for last observation imputed opioid prescription

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Lower CI</th>
<th>Median</th>
<th>Upper CI</th>
<th>psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>20.8</td>
<td>15</td>
<td>-5</td>
<td>19.8</td>
<td>51.9</td>
<td>1.06</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-17.2</td>
<td>11.5</td>
<td>-39.7</td>
<td>-15</td>
<td>0.384</td>
<td>1.02</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-12.1</td>
<td>11.8</td>
<td>-36.9</td>
<td>-9.1</td>
<td>3.95</td>
<td>1.13</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.018</td>
<td>0.093</td>
<td>-0.158</td>
<td>0.014</td>
<td>0.209</td>
<td>1</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>-24</td>
<td>19.4</td>
<td>-62.2</td>
<td>-20.2</td>
<td>4.26</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.004</td>
<td>0.006</td>
<td>-0.017</td>
<td>-0.004</td>
<td>0.006</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_{00}$</td>
<td>106</td>
<td>8.61</td>
<td>91</td>
<td>106</td>
<td>124</td>
<td>1.14</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>0.522</td>
<td>0.397</td>
<td>-0.262</td>
<td>0.552</td>
<td>1.31</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Notes: $\gamma_0$ is the intercept, and $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ are the coefficients for baseline covariates, insurance type, admission type, age of first opioid prescription and pain condition indicator. $\alpha$ is the association parameter and $(\beta_{00}, \beta_{10})'$ is the fixed effect for time.

Table 2.2. Parameter estimates for opioid prescription with missing values

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Lower CI</th>
<th>Median</th>
<th>Upper CI</th>
<th>psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$ (intercept)</td>
<td>-4.47</td>
<td>2.77</td>
<td>-10.2</td>
<td>-4.48</td>
<td>0.872</td>
<td>1.01</td>
</tr>
<tr>
<td>$\gamma_1$ (insurance type)</td>
<td>-0.821</td>
<td>1</td>
<td>-2.49</td>
<td>-0.732</td>
<td>0.848</td>
<td>1.08</td>
</tr>
<tr>
<td>$\gamma_2$ (admission type)</td>
<td>0.717</td>
<td>1.48</td>
<td>-2.35</td>
<td>0.832</td>
<td>3.49</td>
<td>1.01</td>
</tr>
<tr>
<td>$\gamma_3$ (age of first opioid)</td>
<td>-0.056</td>
<td>0.036</td>
<td>-0.132</td>
<td>-0.052</td>
<td>0.008</td>
<td>1.08</td>
</tr>
<tr>
<td>$\gamma_4$ (pain condition indicator)</td>
<td>0.068</td>
<td>1.38</td>
<td>-2.72</td>
<td>0.211</td>
<td>2.71</td>
<td>1.03</td>
</tr>
<tr>
<td>$\alpha$ (association parameter)</td>
<td>0.041</td>
<td>0.021</td>
<td>0.010</td>
<td>0.037</td>
<td>0.082</td>
<td>1.22</td>
</tr>
<tr>
<td>$\beta_{00}$</td>
<td>57.6</td>
<td>3.67</td>
<td>50.7</td>
<td>57.6</td>
<td>64.9</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>0.528</td>
<td>0.171</td>
<td>0.188</td>
<td>0.528</td>
<td>0.862</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: $(\beta_{00}, \beta_{10})'$ is the fixed effect for time.

95% credible interval (0.010, 0.082), indicating a significant positive association between opioid prescription and the onset of OUD.

Then, we removed several covariates with large SDs and only included age of first opioid prescription and pain condition indicator as the covariates in the time-to-event submodel. The shared parameter joint model was fit again with 3 chains of 10,000 MCMC iterations, including 1,000 burn-in iterations. The estimates obtained from this reduced model are shown in Table 2.6.

The potential scale reduction factor (psrf) scores are closer to 1 than those in the full model, indicating a better convergence of the reduced joint model. In the
reduced joint model, the association parameter, \( \alpha \), is still significant. Moreover, the coefficient for age of first opioid prescription, \( \gamma_1 \), is -0.062 with 95% credible interval (-0.114, -0.011), which suggests a significant negative association between the age of first opioid and OUD onset.

### Table 2.3. Parameter estimates for opioid prescription in reduced joint model

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Mean</th>
<th>SD</th>
<th>Lower CI</th>
<th>Median</th>
<th>Upper CI</th>
<th>psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 ) (intercept)</td>
<td>-4.79</td>
<td>2.31</td>
<td>-9.32</td>
<td>-4.79</td>
<td>-0.288</td>
<td>1.02</td>
</tr>
<tr>
<td>( \gamma_1 ) (age of first opioid)</td>
<td>-0.062</td>
<td>0.026</td>
<td>-0.114</td>
<td>-0.061</td>
<td>-0.011</td>
<td>1.01</td>
</tr>
<tr>
<td>( \gamma_2 ) (pain condition indicator)</td>
<td>-0.216</td>
<td>1.44</td>
<td>-3.11</td>
<td>-0.041</td>
<td>2.35</td>
<td>1.01</td>
</tr>
<tr>
<td>( \alpha ) (association parameter)</td>
<td>0.034</td>
<td>0.014</td>
<td>0.009</td>
<td>0.033</td>
<td>0.063</td>
<td>1.02</td>
</tr>
<tr>
<td>( \beta_{00} )</td>
<td>57.4</td>
<td>3.61</td>
<td>50.3</td>
<td>57.3</td>
<td>64.5</td>
<td>1</td>
</tr>
<tr>
<td>( \beta_{10} )</td>
<td>0.532</td>
<td>0.173</td>
<td>0.185</td>
<td>0.532</td>
<td>0.863</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: \((\beta_{00}, \beta_{10})'\) is the fixed effect for time.

### Table 2.4. Comparison of Deviance Information Criteria (DIC)

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint model with imputed values</td>
<td>688.7</td>
</tr>
<tr>
<td>Full joint model</td>
<td>107.9</td>
</tr>
<tr>
<td>Reduced joint model</td>
<td>81.09</td>
</tr>
</tbody>
</table>

The comparison of Deviance Information Criteria (DIC) are shown in Table 2.7, and shows that the reduced model with raw data has the lowest DIC and is more favorable than the other two models. Figure 2.2 - 2.4 provides trace plots for parameters of interest from three models. Comparing the trace plots, we observe that the chains in reduced joint model mixed better than those in the other two models, suggesting a better convergence.

### 2.4 Discussion

In this study, we used a joint model for longitudinal and time-to-event data within the shared parameter framework. As suggested in a previous study (Rizopoulos, 2012), the shared parameter framework allows us to make unbiased estimations without considering the unobserved longitudinal responses. The simu-
Figure 2.2. Trace plots of selected parameters from model with imputed data

imation results show under assumption of MNAR, even when the missing proportion is as large as 40%, the estimates for covariates in the time-to-event submodel are still close to the true values. However, the estimates for the fixed effect in the longitudinal submodel and the association parameter were less satisfactory. In the real data application, the joint model with less covariates performed better than the full joint model and the joint model with imputed longitudinal responses.

The posterior distribution of association parameter, $\alpha$, suggest a weak positive association between longitudinal opioid prescription dosage and the OUD onset. A prior observational study that assessed the prevalence of recurrent opioid usage after an ED opioid prescription revealed that around 30% of all opioid-naive patients prescribed first opioids in EDs went on to engage in recurrent use and were at increased risk for further opioid use one year later (Hoppe et al., 2015). Barnett et al. found among Medicare beneficiaries prescribed a new opioid in the ED, one
out of every 48 patients will become a long-term user (Barnett et al., 2017). Another cross-sectional study observed that among ED patients who reported heroin or prescription opioid abuse and were initially exposed to opioid by a medical prescription, around 30% of them received their first opioids from an ED prescription (Butler et al., 2016). Although these studies suggested opioid prescription in EDs may increase the risk of opioid dependence, the relationship between medical history of opioid prescription and the onset of OUD is still unclear. Our findings show higher values or increasing trends of opioid prescription dosage increase the risk of OUD. Another significant predictor for the time to OUD onset is the age of initial opioid prescription. Younger age has always been supported to be a significant predictor for opioid misuse (Webster, 2017; Turk et al., 2008; Ives et al., 2006; Edlund et al., 2007). Our study focuses on the opioid prescribed from EDs and noted a similar finding. The younger the patient is when he/she gets his/her first opioid prescription from EDs, the greater the risk of OUD becomes.
This study has limitations. First, in the shared parameter joint model, we only considered a shared random effect as association. There are some different association structures for joint model for longitudinal and time-to-event model. For example, the current value association assumes that for individual $y$, the true value of the longitudinal measure at time $t$ is directly related to the risk $h_i(t)$ of experiencing the event at that same time $t$. Then, the joint model can be specified as:

$$h_i(t) = h_0(t)exp(\gamma^T w_i + \alpha \mu_i(t)),$$

where $\mu_i(t) = X_i(t)\beta + Z_i(t)b_i$ is the true value of the longitudinal response. This structure requires the longitudinal submodel to be pre-specified accurately to model the repeatedly observed measurements. Other kinds of association structures also include shared random and fixed effects, current slope, cumulative effects,
etc. Hence, based on the nature of the data, we may consider a different association structure. Second, in this study, we used the Gaussian quadrature method (Rizopoulos, 2012) to approximate the integration in survival function. However, there are some other approaches for the full joint likelihood function, such as adaptive Gauss-Hermite quadrature (Crowther et al., 2012) and Monte-Carlo simulation (Thiébaut et al., 2005). In addition, we can also obtain estimations from the joint model under a Bayesian approach, which we would like to conduct in the future.

2.5 Conclusion

In conclusion, this study conducted a joint analysis of longitudinal and time-to-event data using a shared parameter framework with special consideration of missingness in EHR. A significant association between longitudinal dosage of opioid prescription and the onset of OUD diagnosis is detected in the joint model. Providers should check the history of opioid use and consider a variety of risk factors before prescribing an opioid medication in EDs.

List of References


CDC (2021). Data resources – analyzing opioid prescription data and oral morphine milligram equivalents.


3.1 Introduction

According to the most recent data release, 91,799 Americans died from drug overdoses in 2020, 74.8% of which involved opioids (Hedegaard et al., ). And among all opioid overdoses deaths in 2020, 16,000 deaths, more than 23%, is related to prescription opioids, with over a 16% increase in prescription opioid-involved death rates from 2019 to 2020. (CDC, 2021). On average, 44 people died everyday from overdoses involving prescription opioids. This number increased almost by five times from 1999 to 2020 (CDC, 2021). Despite this alarming trend, the quantity of opioids prescribed per person in terms of morphine milligram equivalents (MME) is still around three times greater than it was in 1999 (Guy Jr et al., 2017). In 2017, for every 100 people in the United States, around 58 people were administered prescription opioids, with an average of 3.4 prescriptions per patient (Hoots et al., 2018).

CDC has been encouraging evidence-based research to identify risk and protective factors and develop effective strategies to prevent opioid-related harms. A key challenge in prevention is the timely identification of people at high risk for opioid dependence or abuse. Common tools used to assess specific opioid-related risk in clinical practice, such as the Opioid Risk Tool (ORT) (Webster and Webster, 2005), the Screener and Opioid Assessment for Patients with Pain (SOAPP) (Akbik et al., 2006), and its revised version (Butler et al., 2009), focus on substance use disorder history, family history of substance use disorder, and significant psychological problems. However, the opioid dependence or abuse is not only associated with social and behavioral factors but
also with physiological and biomedical factors (Ellis et al., 2019; Dong et al., 2019; Katz et al., 2013; Lo-Ciganic et al., 2019). A previous study has shown that the accuracy of the ORT and SOAPP-Revised in predicting subsequent problematic opioid misuse is questionable (Jones and Moore, 2013). To ensure an informed opioid prescribing practice, further effort is needed to develop a comprehensive and reliable opioid-related risk assessment and prediction tool.

With the development of machine learning technology, several prior studies utilized machine learning models to predict opioid-related risks and produced a promising accuracy. Ellis et al. implemented several Random Forest classifiers to predict opioid dependence using lab results and vital signs during 10 or 20 days prior to first substance-related diagnosis (Ellis et al., 2019). With missing data imputation using median or mean, their best classifier can predict whether a patient has substance dependence around 92% of the times. Karhade et al. developed an elastic-net penalized logistic regression to predict continuous opioid prescriptions after total hip arthroplasty (THA) (Karhade et al., 2019b) and achieved a good performance with the area under the receiver operating curve (AUC) of 0.77. However, these studies ignored the temporal patterns within the data (Karhade et al., 2019b; Karhade et al., 2019a; Ellis et al., 2019; Dong et al., 2019). They either measured predictors at a single time point, or treated multiple encounters of one patient as different cases.

Recent studies are attempting to overcome this limitation. Some studies extract some values from fixed time windows, such as the duration, the average value, or the cumulative counts. Lo-Ciganic et al., for example, used cumulative and duration of opioid usage time in 3 months to reflect patient’s opioid use pattern in ML models to predict the risk of OUD in the next 3-month window after initial prescription (Lo-Ciganic et al., 2019; Lo-Ciganic et al., 2020). Even though these
methods allow traditional ML algorithms to be used directly, they have downsides that i) some of the useful temporal information is lost; ii) estimates highly rely on human-assigned duration, requiring extra time and professional help. Deep learning, such as Recurrent Neural Networks (RNN), has been proven very successful in capturing complex temporal relationships among time-dependent features. However, few studies used RNN models with raw temporal data to predict opioid dependence or abuse.

Electronic health records (EHRs) are a valuable resource for developing a deep learning model for predicting opioid-related risk. There are a variety of data types, including features, such as demographic information and disease history, and temporal data during patients’ visits, such as lab test results and outputs over time. However, clinical measurements in EHRs, unlike measurement in experimental studies, are not collected simultaneously at equally spaced time intervals and may not have the same length (Bampa, 2019). In addition, missing observations are common in EHRs due to practical issues. RNN models, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), are incapable of efficiently addressing these irregularly gathered or missing measurements. Additional works are always needed to filter complete cases only or impute missing values using the last available observation, median or MICE. These missing imputation methods may lose valuable information about the time each observation was recorded. Moreover, most missing imputation methods need to be conducted based on the assumption of missing randomness, which is hard to check in EHR data. To deal with this issue, Che et al. proposed GRU-D which modifies the standard GRU by adding a decay rate $\gamma_t$ to models decay mechanism of inputs and hidden output states (Che et al., 2018). GRU-D is built based on two assumptions of missing value mechanism in EHR data: i) the missing values of certain features will become
some default values if the time between last observation and missing observation is long enough and ii) as time passes, the influence of previous observations will decrease. This decay rate can be trained with the model and used to impute missing values at the same time, which saves lots of additional efforts.

However, as we stated before, there are also some static features available in EHR data, which may have a role in the prediction of problematic opioid use, such as history of substance use disorder and age at admission (Amin-Esmaeili et al., 2016; Saha et al., 2016; Edlund et al., 2007). Recurrent Neural Networks (RNN) are beneficial in applications that strongly rely on time-variant variables but may be not capable of taking both static and temporal features as inputs. Most previous studies used RNN models as a feature extractor for temporal features. Activation values of hidden states of the RNN cells are then concatenated with static features as a new set of features that is further fed into a discriminative machine learning model to make the final prediction (Esteban et al., 2016; Leontjeva and Kuzovkin, 2016; Hsu et al., 2019). It is also possible to tackle static feature by transforming them into “fake” sequences that can be fed into RNN models. However, there is a risk of increased loss or decreased accuracy per training time. Considering the structure of GRU-D, if we only include static features in the last step, the missing data imputation will be only based on the temporal feature itself, without considering the inter-correlation between temporal and static features. In EHR, the missingness of historical records may be related to some baseline static features. For instance, a blood urea nitrogen (BUN) test is more like to be conducted if the patient is admitted for a kidney failure or has a history of kidney damage or other chronic condition such as diabetes or high blood pressure. Therefore, by including static information into the GRU-D’s learning process, it may be able to increase prediction accuracy while missing data
and irregular time intervals are present in the data.

Our goal with this work is to predict opioid dependence and abuse during patients’ hospital visits using both static and temporal features. Our predictions are based on patient’s medical history, the sequence of lab test results, the sequence of medications prescribed for the patient, the sequence of input and output charts, as well as patients’ admission information, which includes but is not limited to patients’ age, insurance type, and race. We proposed to build a conditional GRU-D model to embed static features into the initial hidden state as condition variables and evaluate the classification performance of conditional GRU-D model by comparing it with a GRU-D model with only temporal features. In addition, we would like to investigate the feature importance in the best prediction model using Leave One Feature Out (LOFO) Importance technique to detect significant risk factors for opioid dependence and abuse.

3.2 Methodology
3.2.1 Data source and preprocessing

The current study used the Medical Information Mart for Intensive Care III (Mimic-III) (Johnson et al., 2016) provided by the Massachusetts Institute of Technology. This dataset integrates de-identified, comprehensive clinical data of patients admitted to the critical care units of the Beth Israel Deaconess Medical Center in Massachusetts between 2001 and 2012. It includes the information about bedside monitoring, laboratory tests, billing, demographics, diagnoses, and procedures of 58,976 admissions from 48,520 distinct adult patients, with 1190 adult patients having opioid use disorder (OUD) related diagnosis. As the majority of patients only present a single admission, we selected patients who were older than 15 years, had only one admission and stayed in critical care units for at least 48 hours. In order to identify our prediction target – the incidence of OUD,
we used ICD-9 diagnosis codes associated with dependence and abuse of opioids, including 304.00–304.03 (Opioid type dependence), 304.70–304.73 (Combinations of opioids with any other), 305.50–305.53 (Nondependent opioid abuse), 965.0 (Poisoning; Opiates and Related Narcotics), 965.00 (Poisoning; Opium/alkaloids, unspecified), 965.01 (Poisoning; Heroin), 965.02 (Poisoning; Methadone), 965.09 (Poisoning; Other opiates and related narcotics), E85.00 (Accidental Poisoning; Heroin), E85.01 (Accidental Poisoning; Methadone), 97.01 (Poisoning; Opiate antagonists) and E85.02 (Accidental Poisoning; Other Opiates and Related Narcotics) (Heslin et al., 2015). The final dataset included 4534 patients.

Predictors used in our models include 5 categories: personal demographic information (age at admission, race, insurance type etc.), disease history (a set of binary variables, such as history of alcohol use disorder), medication prescriptions during the stay, lab test results, and inputs and outputs records for the patients. From medication prescriptions, we selected 27 drugs that were prescribed most often. For lab test results, inputs, and outputs charts, we selected the top 20 items with most occurrence from each. Therefore, we included 87 temporal predictors in our models. All continuous features were standardized and values that are 4 standard deviation away from the mean and abnormal values were treated as outliers and removed. Categorical features were dummy coded into several binary features so we included 26 static predictors in our models.

3.2.2 Gated Recurrent Unit-D Model

Recurrent Neural Network (RNN) is a type of DL models for analyzing temporal data (Goller and Kuchler, 1996). RNN models perform the same action at each time step of the sequence input and feed the output as part of the input to the next time step, which enables RNN to memorize and update parameters through all time steps. Among RNN models, Long Short-Term Memory (LSTM)
and Gated Recurrent Unit (GRU) are most widely used. GRU was first introduced by Cho et al. in order to make each recurrent unit to adaptively capture dependencies of different time scales (Cho et al., 2014). GRU can be implemented as follow (Chung et al., 2014):

\begin{align*}
\text{Update state} : z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (3.1) \\
\text{Reset gate} : r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (3.2) \\
\text{Hidden state : activation} : h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t; \quad (3.3) \\
\text{candidate activation} : \tilde{h}_t &= \tanh(W x_t + U (r_t \odot h_{t-1}) + b), \quad (3.4)
\end{align*}

where \( t \) is the time step, \( \sigma \) is the logistic sigmoid function, and \( \odot \) represents element-wise product. As shown in Figure 3, at each time step \( t \), GRU has a reset gate \( r_t \) and an update gate \( z_t \) for each of the hidden state \( h_t \), which is treated as the output of the input \( x_t \). Then, we smooth \( h_t \) at every time step and apply another sigmoid layer over them to make the prediction \( y \).

GRU have been shown to achieve state-of-art performance for handling temporal data. However, in EHR data, most intervals between two measurements are not evenly spaced, and measurements are not collected simultaneously and may not have the same length. These irregular sampled data cannot be modeled effectively by GRU (Bang et al., 2020; Neil et al., 2016; Che et al., 2018). Thus, Che et al. proposed GRU-D, which adds decay rates \( \gamma_t = \exp\{-\max (0, W_\gamma \delta_t + b_\gamma)\} \) to standard GRU to model the decay mechanism of inputs and hidden output states (Che et al., 2018). \( \delta_t \) is a matrix of the time interval between two observations. Missing observations will be imputed with \( \hat{x}_t^d = m_t^d x_t^d + (1 - m_t^d)(\gamma_t^d x_t^d + (1 - \gamma_t^d) \hat{x}_t) \). \( m_t^d \) is a masking variable, which is 0 when the current observation is missing, 1 if observed. The author used GRU-D in two health care applications of mortality prediction and found GRU-D outperformed basic GRU models.
3.2.3 Combining static and temporal features

As discussed before, EHR data contain both temporal information and static information. Some static variables, such as age at admission, gender, race, and disease history, are significantly associated with problematic opioid use. Some previous studies suggest being male is a positive predictor for opioid misuse and abuse (Edlund et al., 2007; Ives et al., 2006). Younger age is also supported to be a significant predictor for opioid misuse (Webster, 2017; Turk et al., 2008; Ives et al., 2006; Edlund et al., 2007). In addition, a personal history of non-opioid substance abuse was found to be one of the most consistent predictors for opioid misuse among chronic pain patients (Turk et al., 2008).

Therefore, we build a GRU-D model that can include both static features and temporal features in OUD prediction. The first and simplest approach is to include all static features at the first time step so that these features can influence the hidden state of the GRU-D model. But RNN models are not able to memorize data for long time and may forget its previous inputs, especially input at first time step. So this approach may be not able to model the influence of static features when the sequence of temporal features are long.

Van den Oord et al. introduced a conditional Pixel convolutional neural network (CNN) to condition PixelCNN models for images on any vector, including descriptive labels or tags, or latent embeddings created by other networks.
(Van den Oord et al., 2016). Their experiments show that with additional condition vectors, conditional PixelCNN can generate more realistic looking images than the original PixelCNN. Based on this idea, we propose a conditional GRU-D model that can condition temporal features on static features. In conditional GRU-D model, static features are treated as condition variables $x$ to initiate the hidden state of the GRU-D. We follow the steps described as below:

1. For each training sample, take static features as a condition variable vector $\vec{x}$

2. Transform condition variable vector $\vec{x}$ with an affine transformation to get it into the same shape as the hidden state of the RNN, $\vec{v} = W\vec{x} + \vec{b}$, where $W$ and $\vec{b}$ are trainable weights. This step is achieved using a linear transformation in Pytorch (Paszke et al., 2019).

3. In the initial step, add $\vec{v}$ to the hidden state of the GRU-D when computing its value.

Follow these step, GRU-D model with temporal inputs can be properly conditioned on static inputs.

### 3.2.4 Model Development and Evaluation

We also considered a GRU-D model with only temporal features as the baseline model, as well as a GRU-D model with all static features included at the first time step. The models were implemented in Python using Pytorch (Paszke et al., 2019).

We randomly split the data into three subsets: training set, validation set, and test set. The training set included 60% of the patients, 20% were included in the validation set, and another 20% of the patients were assigned to the test set. Different experiments were performed to compare the performance of three models,
with different sets of hyperparameters, such as learning rate and number of hidden units. For optimization, we chose the Adam algorithm (Kingma and Ba, 2014) to update the weights since it is currently recommended as the default gradient descent optimization algorithm to use (Ruder, 2016). Binary Cross Entropy between the target and the input probabilities was used to estimate the prediction loss. We repeated each experiment five times with different random splits of the data.

The performance of the models was evaluated and compared using the mean area under the Precision-Recall curve (AUPRC), mean area under Receiver Operating Characteristics curve (AUROC), and F1-score. Since the target group we want to predict – OUD patients only accounts for a small portion of the total sample, we focused more on detecting when an opioid abuse or dependence is going to happen rather than when it is not going to happen. We would like to get a high sensitivity value (true positive rate) and focus less on specificity (true negative rate). Precision is the percentage of correct positive predictions relative to total positive predictions, and recall is the percentage of correct positive predictions relative to total actual positives. Precision-recall (PR) curves plot the positive predictions against the true positive rate, and F1-score is a weighted harmonic mean of precision and recall. Thus, PR curves and F1-score are a better metrics to evaluate the prediction performance when the target event happens much less frequently.

3.2.5 Feature Importance

GRU-D models, like other neural networks, are black box models that aim for predictive accuracy rather than for inference. It’s hard to understand the role of each predictor in the prediction model. Leave-One-Covariate-Out (LOCO) technique was proposed by Lei et al. (Lei et al., 2018). This approach compares the performance of a whole model with the performance of a model which is fitted with
a subset of the features. We implemented this technique using Pytorch following the steps below:

1. Train an original model and estimate a baseline model loss $loss_0$, such as binary cross entropy loss in this study

2. For each feature $j = 1, \ldots, p$, remove feature $j$ from the data $X$ and refit model with data $X_{-j}$

3. Estimate loss based on the predictions of the reduced data $X_{-j}$ and Calculate LOCO Feature Importance $FI_j = loss_{-j} - loss_0$

3.3 Results

3.3.1 Descriptive statistics and missingness

Among 4534 patients aged 16 to 107 years in the study sample, 15.04% ($n = 682$) had opioid dependence or abuse related diagnosis. Their demographic collected at admission are reported in Table 6. Most of the patients were white (70.58%) and admitted from emergency department (86.35%). There were more patients with a history of hypersensitivity lung disease than those with a history of anxiety disorder or alcohol use disorder. More than half of the patients were covered by Medicare (55.56%).

In the dataset, both the irregular sampling issue and the asynchronism of feature sampling problem were observed. The 87 temporal features were collected at varying frequency across the data. Some variables, such as blood test results, were measured far more often than others, whilst others were measured seldom. We picked a feature at random from the medication prescriptions, lab test results, and inputs and outputs records. As shown in Figure 4, urine void from the output charts and the morphine sulfate drug had around 550 counts while chloride blood test had more than 3,500 records. The use of the GRU-D model was encouraged to address this problem to predict OUD with temporal features from EHR data.
3.3.2 Model Performance

Three models were trained and validated on 5 different sets of training and validation sets. Table 7 shows the performance results of three models in predicting opioid dependence or abuse in the test sets, including AUROC, AUPRC, and F1-score. In terms of AUROC, the GRU-D model which included static features at the first time step was the best-performed model (0.793) followed by conditional GRU-D model (0.676), and GRU-D with temporal features only (0.658). All three models’ AUROCs are higher than 0.5, which means the predictions made by the models are better than a random guess. Same findings can also be observed from Figure 5. The ROC for GRU-D with static features added at the first step is farther away from the no-skill diagonal toward the upper-left corner than the curves for the other two models. Based on AUPRC, GRU-D with static features added at the first time step still has the best performance with much a larger mean AUPRC value than the other two models. The conditional GRU-D’s AUPRC is slightly larger than the one for GRU-D with only temporal features. Figure 6 shows that the curve for GRU-D with static features added at the first time step bows more towards (1,1) above the flat x-axis of no skill than those of conditional GRU-D and GRU-D with only temporal features. F1-scores also show a similar pattern, with GRU-D with static features added at the first time step the best-performed model (0.38).
Overall, the GRU-D with static features added at the first time step has the best prediction performance. In addition, based on recall scores, out of all the patients that actually did get an opioid dependence or abuse related diagnosis, this model predicted this outcome correctly for 29% of those patients. Based on precision score, 53% of all the patients that the model predicted would get an opioid dependence or abuse related diagnosis actually got opioid dependence or abuse related diagnosis.

Figure 3.3. Model performance based on the area under the receiver operating characteristic curve (AUROC)

### 3.3.3 Risk factors based on feature importance

We evaluated the importance of 113 features in the best-performed model, GRU-D model with static features added at first time step, using Leave-One-Covariate-Out (LOCO) technique. We selected the top 15 features that have the highest feature importance scores and plotted their average loss against the baseline loss in Figure 7. The distance from the baseline loss line to the right edge of the
Glucose blood test result was ranked to be the most important predictor for opioid dependence or abuse. Among the 15 most important features, 4 are static features, including being a Native American, being widowed, having a history of anxiety disorder, and being covered by Medicare. Medications, like magnesium sulfate drug, Lactated Ringer’s injection, ondansetron drug, morphine sulfate, also play a role in the prediction of opioid use disorder.

3.4 Discussion

This study yield valuable insights because of the use of RNN to develop a prediction model for opioid dependence and abuse based on static and temporal features from the EHR simultaneously. Many machine learning studies that predict opioid-related risks ignore the temporal structure within the EHR data (Ellis et al., 2019; Karhade et al., 2019b; Dong et al., 2019). As a result of the
growth of deep learning (DL) models, an increasing number of research papers are employing DL with EHR data to forecast a range of health-related risks in a long term, including in-hospital mortality, diagnoses classification, and length of stay (Lipton et al., 2015; Rajkomar et al., 2018; Bang et al., 2020). However, few studies have used DL models in opioid related predictions or combined static features in the prediction. In the current study, we built and compared several prediction models based on GRU that not only can handle irregular sampling issue but also are capable of combining both static and temporal variables from EHR data, in order to solve the task of predicting opioid dependence or abuse among patients in critical care units. The comparison of model performance suggests adding static features as condition variables can slightly improve prediction performance of the GRU-D model only based on temporal features. Our findings indicate that simply integrating static features at the first time step in the GRU-D model may result in a reasonably good prediction, even better than the prediction made by conditional GRU-D model. We hypothesize that this may be due to the insufficient size of the sample used in this study. Therefore, the best-performed model has 27 more features than the conditional GRU-D model due to the addition of 27 static features at the first time step.

In addition, we investigated the feature importance in the best-performed model using LOCO approach and the results show that static features, including, being a Native American, being widowed, having a history of anxiety disorder, and being covered by Medicare are important factors to predict opioid dependence or abuse. Anxiety disorder or mood disorder has been widely recognized as a significant risk factor of OUD (Martins et al., 2012; Vorspan et al., 2015; Blanco et al., 2013; Martel et al., 2013; Fischer et al., 2012). Martins et al. conducted a national longitudinal study using 34,653 adults’ surveys from the Na-
tional Epidemiologic Survey on Alcohol and Related Conditions and found that baseline mood disorders were significantly associated with incident opioid disorder due to prescription opioid use at follow-up (Martins et al., 2012). Vorspan et al. conducted a literature review on the association of any substance use disorders and anxiety (Vorspan et al., 2015). Among different kinds of substance use disorder, anxiety disorders or anxiety symptoms are constantly found significant in multiple studies that investigated factors associated with prescription opioid dependence (Blanco et al., 2013; Martel et al., 2013; Fischer et al., 2012). Another notable predictor of opioid use disorder is insurance type. Medicare is the federal health insurance program for people aged 65 and older, certain younger people with disabilities and people with End-Stage Renal Disease. So being covered by Medicare may indicate that this patient is older than 65 years old, or having certain disability or severe renal disease. Younger age is supported to be a significant predictor for opioid misuse in lots of previous studies (Webster, 2017; Turk et al., 2008; Ives et al., 2006; Edlund et al., 2007). In term of renal disease, chronic pain is a common symptom among patients with end-stage renal disease and there is a high use of opioid among this population (Roy et al., 2020). Thus, they may be exposed to higher risk of opioid dependence. In addition to static predictors, temporal predictor, glucose blood test result is identified as the most important predictor in our GRU-D model. Glucose test is more likely to be prescribed to patients with diabetes, and opioids are often prescribed to manage diabetes-related neuropathic pain (Jensen et al., 2006). Long-term use of opioids may lead to opioid misuse and abuse (Walter et al., 2017). In addition, evidence from pre-clinical and clinical studies suggest that opioid use is associated with glucose dysregulation, which may let the patient need a glucose test (Mysels and Sullivan, 2010). Among all the drug related important features, besides morphine sulfate, mag-
nesium sulfate drug and ondansetron drug are also found associated with opioid dependence and abuse. A clinical study demonstrates that magnesium may reduce the severity of opioid addiction and significantly relieve withdrawal syndrome symptoms (Nechifor et al., 2004). Moreover, according to a prior research, treatment with the 5-HT3 receptor antagonist (5-HT3-RA) ondansetron may decrease objective measurement of Opioid withdrawal (OW) in adults by up to 76% (Chu et al., 2009). The intensity of OW symptoms is a primary factor in the risk for dependence to prescribed and illicit opioids. Ondansetron was also found to be able to decrease alcohol consumption and cravings in patients with alcohol use disorder (Kranzler et al., 2003). Hence, ondansetron drug’s effect on opioid addition is worth continuing to investigate.

This study has limitations. First, to combine static and temporal features, we modified the structure of GRU-D model, or just added static features at the first time step. But the results suggest that the performance of our proposed conditional GRU-D model performed was not as good as the simple GRU-D model. Secondly, GRU-D model, like most neural networks, is a black-box model. Even though we estimated the importance of features in the prediction using LOCO approach, we cannot know whether this feature is a protective factor that is associated with a lower likelihood of opioid dependence and abuse, or a risk factor that is associated with a higher likelihood of our target events. Thirdly, this study only consists of around 4,000 patients. To develop a practical decision support model, we need to include as much patients as possible to make reliable predictions for future events.

For next steps, we first would like to develop an ensembled GRU-D model. Recently, there are studies exploring how to handle both static and temporal features using ensembled RNN model (Esteban et al., 2016). In the ensembled RNN model, RNN is used as a temporal feature extractor, and the activations of the last
neural network layer at the last iteration are combined with the static features as a new set of features. Then a traditional machine learning model, such as random forest (RF) model, can be trained on this new set of features to make predictions. Second, to investigate whether a feature is a protective or risk factor, there are several methods we can utilize in the future. For example, we can transform the values of a feature to negative values to fit the model and then set the values to positive values to fit the model again. So we can investigate how the average effect of this feature on target variable changes. Another approach is to split the whole sample to subgroups based on certain feature in order to compare the effect of this feature on target variable in different subgroups. Lastly, we plan to integrate more data into our model to improve the performance of our prediction for opioid dependence and abuse. In addition, due to the limited sample size, we only included temporal features from medication prescription, lab test results, and input and output charts. With a larger sample size, we may be able to incorporate more information from EHR data, such as vital signs like heart rate. Last but not least, since more than 70% of the patients in this study are white patients, the fairness of this prediction model needs further investigation. This topic has been increasingly discussed in Artificial Intelligence (AI) applications, especially in some sensitive areas, such as health care and criminal justice. AI models are driven by data. Hence, when a model is trained on a data set, where a minority group is not sufficiently represented, the results produced may be misleading and the algorithm may be unfair (Veale and Binns, 2017). Hence, a prediction model based on a more diverse group will be developed in the future.

3.5 Conclusion

This study demonstrates the feasibility and potential of GRU-D models to predict opioid dependence and abuse using static and temporal information from
EHR data. Conditional GRU-D model and GRU-D model with static features added at first time step can produce a better prediction than the model only based on temporal features. Especially, the GRU-D model with static features added at first time step has fairly good prediction performance and appear to be a valuable tool to identify patients at high risk of opioid dependence and abuse. This study also investigated and revealed the importance of several features from laboratory tests and medication prescriptions in the prediction of problematic opioid use.

List of References


<table>
<thead>
<tr>
<th>Feature</th>
<th>% or Mean (SD)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at Admission</td>
<td>70.58 (50.43)</td>
<td>4534</td>
</tr>
<tr>
<td>Admission Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective</td>
<td>11.84</td>
<td>537</td>
</tr>
<tr>
<td>Emergency</td>
<td>86.35</td>
<td>3915</td>
</tr>
<tr>
<td>Urgent</td>
<td>1.81</td>
<td>82</td>
</tr>
<tr>
<td>Insurance Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government</td>
<td>3.13</td>
<td>142</td>
</tr>
<tr>
<td>Medicaid</td>
<td>14.20</td>
<td>644</td>
</tr>
<tr>
<td>Medicare</td>
<td>55.56</td>
<td>2519</td>
</tr>
<tr>
<td>Private</td>
<td>26.29</td>
<td>1192</td>
</tr>
<tr>
<td>Self pay</td>
<td>0.82</td>
<td>37</td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>7.96</td>
<td>361</td>
</tr>
<tr>
<td>Life partner</td>
<td>0.04</td>
<td>2</td>
</tr>
<tr>
<td>Married</td>
<td>43.91</td>
<td>1991</td>
</tr>
<tr>
<td>Separated</td>
<td>1.21</td>
<td>55</td>
</tr>
<tr>
<td>Single</td>
<td>30.68</td>
<td>1391</td>
</tr>
<tr>
<td>Widowed</td>
<td>12.55</td>
<td>569</td>
</tr>
<tr>
<td>Unknown</td>
<td>0.49</td>
<td>22</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>70.58</td>
<td>3200</td>
</tr>
<tr>
<td>Black or African American</td>
<td>13.98</td>
<td>634</td>
</tr>
<tr>
<td>Asian</td>
<td>2.25</td>
<td>102</td>
</tr>
<tr>
<td>Native American</td>
<td>0.15</td>
<td>7</td>
</tr>
<tr>
<td>Unknown/not specified</td>
<td>12.73</td>
<td>591</td>
</tr>
<tr>
<td>Anxiety Disorder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3.11</td>
<td>141</td>
</tr>
<tr>
<td>No</td>
<td>96.89</td>
<td>4393</td>
</tr>
<tr>
<td>Alcohol Use Disorder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2.30</td>
<td>104</td>
</tr>
<tr>
<td>No</td>
<td>97.70</td>
<td>4430</td>
</tr>
<tr>
<td>Hypersensitivity Lung Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15.01</td>
<td>680</td>
</tr>
<tr>
<td>No</td>
<td>85.00</td>
<td>3854</td>
</tr>
</tbody>
</table>
Table 3.2. Model performance based on mean AUROC, AUPRC, and F1-score

<table>
<thead>
<tr>
<th>Model Description</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional GRU-D</td>
<td>0.676</td>
<td>0.281</td>
<td>0.09</td>
</tr>
<tr>
<td>GRU-D</td>
<td>0.793</td>
<td>0.417</td>
<td>0.38</td>
</tr>
<tr>
<td>with static added at 1st time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRU-D</td>
<td>0.658</td>
<td>0.273</td>
<td>0.06</td>
</tr>
<tr>
<td>with temporal only</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.5. Importance of the features in the best-performed model (GRU-D model with static added at first time step) based on LOCO
APPENDIX

Code for shared parameter model in Manuscript 2

A.1 R code for shared parameter joint model in JAGS

```r
model_string <- "

model{
for (i in 1:n ){
    # Longitudinal submodel
    for ( j in 1 :M[i]){
        Yij[i,j] ~ dnorm(mu[i,j],tau)
        mu[i,j] <- inprod(betaL[], XL[i,j, ] ) + inprod(b[i,],ZL[i,j, ])
    }
    # Hazard function
    for( j in 1:K){
        haz[i,j] <- lambda*pow(Time[i]/2 * (xk[j]+1), lambda - 1)*
            exp(inprod(betaS[], XS[i, ])+
                alpha*(b[i,1] + b[i,2] *(Time[i]/ 2*(xk[j] + 1 ))))
    }
    # Log-survival function with Gauss-Legendre quadrature
    logSurv[i] <- -Time [i]/2 * inprod(wk,haz[i ,])
    # survival log-likelihood using zero tricks
    phi[i] <- 100000 - death[i] * log(haz[i,K]) - logSurv[i]
    zeros[i] ~ dpois(phi[i])
    # Random effects
    b[i,1:Nb] ~ dmnorm(mub[],Omega[ ,])
}
    # Priors
    for(l in 1:NbetasS){ betaS[l] ~ dnorm(0.0, 0.001) }
    alpha ~ dnorm(0.0, 0.001)
    lambda ~ dunif(0, 10)
    for(l in 1:NbetasL){ betaL[ l ] ~ dnorm(0.0, 0.001) }
    tau <- pow(sigma , -2)
    sigma~ dunif(0, 100)
"
```

```
\text{Omega}[1: Nb, 1: Nb] \sim \text{dwish}(V[,], N_b) \\
\text{Sigma}[1: Nb, 1: Nb] \leftarrow \text{inverse}(\text{Omega}[,])

\# Derived quantity

\text{q} \leftarrow \exp(\text{betaS}[1])
}

"
APPENDIX

Simulation for shared parameter model with missing time-varying covariates in Manuscript 2

B.1 Simulation

The performance of the shared parameter joint model regarding the percentage of missing values is evaluated using 100 simulated datasets with each having $N = 100$ individuals. For simplicity, we include one binary time-independent covariate $x_{1i}$, one continuous time-independent covariate $x_{2i}$ in time-to-event sub-model, and 1 time-varying longitudinal variable $y_i(t)$. Three scenarios with different percentage of missing were considered, including no missing, 20% missing and 40% missing. And the missing values in $y_i(t)$ were assumed MAR, which means the probability of missing depends on observed time-independent covariates $x_{1i}$ and $x_{2i}$. The complete datasets were simulated using simjm package (Brilleman, 2018) following the forms below:

$$y_i(t) = \beta_{00} + b_{0i} + (\beta_{10} + b_{1i})t + \epsilon_i(t)$$
$$h_i(t) = \lambda t^{\lambda - 1} \exp(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \alpha (b_{0i} + b_{1i})t),$$

where $\beta_{00} = 2, \beta_{10} = -0.5, \gamma_0 = -3.3, \gamma_1 = -0.5, \gamma_2 = 0.5$, and $\alpha = 0.5$. The binary covariate $x_{1i}$ was simulated from $Bern(0.5)$ and the continuous covariate $x_{2i}$ was generated from a normal distribution $N(0, 1)$. The random effects $(b_{0i}, b_{1i})'$ satisfy a normal distribution with mean of 0 and variance of $\Sigma$. The variance-covariance matrix, $\Sigma$, can be represented as $\Sigma = VRV$, where $R$ is the correlation matrix and $V = diag(\sigma_b)$ is a square diagonal matrix with the standard deviations as the diagonal elements. We consider $\sigma_b = (2, 1)$ and $R = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. The R code is included in Appendix A.
Each simulated dataset were then processed to create two more datasets with 20% or 40% missing values in \( y_i(t) \). We fit the shared parameter joint model for each simulated dataset using 3 chains of 10,000 MCMC iterations, which included 5,000 burn-in iterations that were not used for inference.

B.2 R code for simulating longitudinal and time-to-event data using \texttt{simjm} function

```r
dat <- simjm(n = n, M = 1, fixed_trajectory = "linear",
             random_trajectory = "linear",
             assoc = "shared_b(1)",
             basehaz = c("weibull"),
             betaLong_intercept = 2,
             betaLong_binary = 0,
             betaLong_continuous = 0,
             b_sd=c(2,1),b_rho=0.5,
             betaLong_linear = -0.5,
             betaEvent_intercept = -3.3,
             betaEvent_binary = -0.5,
             betaEvent_continuous = 0.5,
             betaEvent_assoc = 0.5,
             max_yobs = 10,
             max_fuptime = 10, balanced = TRUE,family = gaussian)
```

B.3 Simulation Results

To compare the model performance under different missing percentages, we calculated and reported the following estimates using 100 simulated datasets:

- the mean of the posterior distribution for each parameter
- the mean standard deviation (SD) of the posterior distribution for each parameter
- the standard error (SE) of the posterior mean for each parameter, defined as the standard deviation of the estimates
- the 95% credible interval of the posterior distribution form each parameter.
Table B.1. Parameter estimates for complete dataset

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>-3.486</td>
<td>0.439</td>
<td>0.927</td>
<td>-4.352</td>
<td>-2.638</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.679</td>
<td>0.301</td>
<td>0.818</td>
<td>-1.269</td>
<td>-0.089</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.666</td>
<td>0.172</td>
<td>0.459</td>
<td>0.330</td>
<td>1.005</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.001</td>
<td>0.021</td>
<td>0.020</td>
<td>-0.040</td>
<td>0.042</td>
</tr>
<tr>
<td>$\beta_{00}$</td>
<td>2.022</td>
<td>0.193</td>
<td>0.207</td>
<td>1.645</td>
<td>2.396</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>-0.492</td>
<td>0.099</td>
<td>0.110</td>
<td>-0.681</td>
<td>-0.305</td>
</tr>
</tbody>
</table>

Table B.2. Parameter estimates for dataset with 20% missing values (MNAR)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>-3.492</td>
<td>0.440</td>
<td>0.929</td>
<td>-4.360</td>
<td>-2.642</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.679</td>
<td>0.302</td>
<td>0.821</td>
<td>-1.270</td>
<td>-0.088</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.669</td>
<td>0.173</td>
<td>0.463</td>
<td>0.332</td>
<td>1.011</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.002</td>
<td>0.031</td>
<td>0.027</td>
<td>-0.061</td>
<td>0.059</td>
</tr>
<tr>
<td>$\beta_{00}$</td>
<td>1.299</td>
<td>0.197</td>
<td>0.180</td>
<td>0.911</td>
<td>1.684</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>-0.788</td>
<td>0.105</td>
<td>0.118</td>
<td>-0.987</td>
<td>-0.587</td>
</tr>
</tbody>
</table>

From Table 2.1 - 2.3, we observe the estimates for intercept and covariates’ coefficients in time-to-event submodel are similar across three scenarios. As the missing percentage increases, the difference between true value and estimated value increases slightly, and so do the SD and SE. However, the posterior mean for association parameter $\alpha$ are close 0, differing significantly from the true value. In addition, there is a large difference between the 3 posterior means. We suspect this may be related to the estimates of intercept and slope in the longitudinal submodel. Because the posterior mean for fixed-effect intercept and slope gets far from the true value when the missing percentage increases.

The simulation results didn’t meet our expectations, especially for the estimates of association parameters. We suspect this might be related to the R package we used to simulate joint longitudinal and time-to-event data. Although the author of this package showed $simjm$ function can simulate distributions close to the target distribution (Brilleman, 2018), the whole simulation process is not transparent. We should consider using a different package, such as $simsurv$ (Brilleman et al., 2020),
Table B.3. Parameter estimates for dataset with 40% missing values (MNAR)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>-3.494</td>
<td>0.440</td>
<td>0.938</td>
<td>-4.361</td>
<td>-2.644</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.682</td>
<td>0.303</td>
<td>0.825</td>
<td>-1.278</td>
<td>-0.091</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.670</td>
<td>0.174</td>
<td>0.464</td>
<td>0.332</td>
<td>1.012</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.003</td>
<td>0.034</td>
<td>0.031</td>
<td>-0.064</td>
<td>0.070</td>
</tr>
<tr>
<td>$\beta_{00}$</td>
<td>1.048</td>
<td>0.212</td>
<td>0.199</td>
<td>0.635</td>
<td>1.465</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>-0.869</td>
<td>0.107</td>
<td>0.123</td>
<td>-1.073</td>
<td>-0.659</td>
</tr>
</tbody>
</table>

or simulating data from sketch to make sure all the parameters are specified correctly.

**List of References**


CDC and NCHS (2020). Wide-ranging online data for epidemiologic research (wonder).


SAMHSA (2020). Key substance use and mental health indicators in the United States: Results from the 2019 national survey on drug use and health (hhs publication no. pep20-07-01-001, nsduh series h-55). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.


CDC (2021). Data resources – analyzing opioid prescription data and oral morphine milligram equivalents.


Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*, pages 475–492. NBER.


Thiébaut, R., Jacqmin-Gadda, H., Babiker, A., Commenges, D., and Collabora-


